TT 715
.M 8

Columbia University

Contributions to Education

Teachers College Series

Class TT715
Book .M8
Copyright No.

COPYRIGHT DEPOSIT.

The Measurement of Certain Elements of Hand Sewing

By Katharine Murdoch, Ph.D.

Teachers College, Columbia University
Contributions to Education, No. 103

Published by
Teachers College, Columbia University
New York City
1919

Monograph

Copyright, 1919, by KATHARINE MURDOCH

TT715
.M8

MAR 18 1920

©CI.A566539

CONTENTS

I. The Problem 1

II. The Data 8

Subjects
Material
Judges

III. The Scale 13

IV. Faults in Sewing 32

The Analysis of Faults
The Measurement and Significance of Various Faults

V. The Reliability of Measurements of Sewing . . . 90

Reliability in Relation to the Kind of Stitch
Reliability in Relation to Synthetic *vs.* Analytic Judgments
Reliability in Relation to the Number of Samples
Reliability in Relation to the Character of the Judges
Reliability of Judgment in Relation to the Distribution and to the Kind of Products Judged
Reliability in Relation to the Use of the Scale

VI. A Summary of Results 108

Appendix I. A Test of Transfer from Sewing to "Neatness" and to "Æsthetic Appreciation" 115

Appendix II. The Scale . . . . . 120 *and Insert*

ACKNOWLEDGMENTS

Without the very kind coöperation of many persons this research could not possibly have been carried through. The author is under a great debt of gratitude to those who assisted her in procuring the sewing samples used. Miss K. Antes, of Port Chester, N. Y.; Mr. H. H. Stewart, of Mount Vernon, N. Y.; Miss Hinman, of the Carnegie Institute of Technology; Miss Vogelsang, of Kingsley House, Pittsburgh, Pa.; Miss Fawcett, of East Orange, N.J.; Superintendent D. C. Bliss, of Montclair, N. J.; Mrs. Bounty, of Stamford, Conn.; Miss Spohr, of Teachers College; Miss Sallie Fisk and Miss Luella Smock are among those who assisted in this way.

To all the busy persons who made judgments upon the samplers, and to the many who gave helpful advice and assistance of various sorts, the author is indeed grateful. Among these latter she particularly feels indebted to Mrs. Consaulus, of the Washington Irving High School; Professor Cooley, of Teachers College; Miss Ethelwyn Miller, of the Iowa State College; Dr. Kate Gordon, of the Carnegie Institute of Technology; Miss Mary Ingersoll, of Cornell University, and Miss Meda C. Lynn.

Dr. William A. McCall and Dr. Truman Lee Kelley, of Teachers College, were very helpful in the giving of advice and suggestions concerning statistical matters.

Above all, the author desires to express her sense of indebtedness to Dr. E. L. Thorndike, without whose suggestions, guidance, and inspiring influence, given at a time when he was under unusual strain, this work would never have been accomplished.

K. M.

SECTION I

THE PROBLEM

Hand sewing, with its predecessor in garment making, weaving upon the loom, is one of the oldest of the arts, and the teaching of it long antedates our schools. Penelope in her upper chamber, bidding "her women ply their tasks at the distaff and loom," was far from a pioneer in the teaching of woman's first art. A history of this teaching would form a large share of the total history of the education of women before the time of their formal instruction in schools, and runs parallel with this history from then on. The early records of Boston show that needlework was taught to girls by their regular teachers as early as 1798, when they were first admitted to the schools. Not until 1873, however, was a regular sewing teacher appointed. From that time the teaching of sewing as a specialized element of school instruction has gradually spread, thus somewhat preceding the more general movements of manual training and industrial arts education.

In spite of its long history and wide use, sewing has received very little attention from the psychologist or from efficiency engineers or measurers of educational products. We still use phrases like the "good work" and "handsomest in work" by which Homer measured the lovely robe of Helen of Troy. A somewhat more objective system of measurement was introduced when, with the coming of school grades upon a percentage or letter basis in other subjects, came also the same system for grading the sewing work in a school. Both the adjective method and the school-grade method are still in vogue and await the coming of more objective measures.

In a few of the western universities and agricultural schools there has been put in use rather recently score cards to measure sewing products. These cards indicate the maximum per cent that can be awarded for various parts of the work, and refer sometimes to the process of sewing as well as to the products—sometimes to the product alone. Below are two samples of such score cards:[1]

[1] These score cards were devised by Miss Katherine Cranor, of the University of Nebraska.

SCORE CARD FOR UNDERWEAR

I. Technique 45%
- *a.* Cutting
- *b.* Fitting
- *c.* Workmanship

II. Selection of material . . . 20%
- *a.* Quality
- *b.* Suitability
- *c.* 1. Garment
  2. Use
- *d.* Relation of trimming to material of garment
- *e.* Hygiene—laundering qualities

III. Design 15%
- *a.* Good spacing
  1. Hems
  2. Tucks
  3. Embroidery and other trimmings
- *b.* Originality

IV. Neatness 15%
- *a.* Clean
- *b.* Finishings
- *c.* Well pressed

V. Cost 5%
- *a.* Value of money expended
- *b.* Economy in trimming

SCORE CARD FOR DRESSMAKING

I. Design 25%
- *a.* Line
- *b.* Color
- *c.* Suitability as to type and occasion
- *d.* Silhouette
- *e.* Originality and individuality

II. Selection of material . . . 15%
- *a.* Color
- *b.* Suitability as to use and age
- *c.* Trimming
- *d.* Durability

III. Technique 25%
- *a.* Cutting
- *b.* Fitting
- *c.* Workmanship
- *d.* Pressing
- *e.* Neatness

IV. Hygiene 15%
- *a.* Cut and construction
- *b.* Material
- *c.* Cleaning qualities

V. Cost 10%
- *a.* Value of money expended
- *b.* Standard of living
- *c.* Relationship of trimming to material

VI. Ethics 10%
- *a.* Modesty
- *b.* Influence

These score cards vary and depend for their usefulness upon the judgment of their makers as to what really is the relative importance of each part in relation to the whole.

There are few, if any, elementary school subjects to-day which have received as little serious study on the measurement side as has this ancient art of hand sewing. Because no means of scientific measurement for this subject is at hand, it has been impossible to

draw any well-founded scientific conclusions as to many important questions of methods and results.

The present study grew out of a desire upon the author's part to evaluate scientifically certain methods of teaching sewing. Only a very few of the questions which led to this research are here solved or, for that matter, even attacked. The author hopes, however, that besides answering a few of her original questions, this investigation will furnish a measuring instrument for certain parts of the whole subject of elementary sewing which will be freely used by others who may make investigations in this field and so free their effort for the attack upon more interesting problems than have been studied here.

An almost indefinite number of these problems, all of them requiring some means of measurement for merit of the finished product of sewing, crowd to mind. I shall mention only a very few which will be quickly added to by every progressive teacher of sewing.

There are, first, many problems of wide concern to educators in general, such as the amount of transfer which can be expected from the teaching of sewing to other fields, in ideals of workmanship, creative and æsthetic ability, etc. There are also general problems of method in the teaching of sewing itself, such as: (1) individual *versus* group instruction; (2) the amount of transfer to be expected from knitting, card-sewing, and the like, to ability to sew on cloth; (3) the so-called "logical" method of starting with elementary tasks, such as stitches upon samplers *versus* the so-called "psychological" method of starting with some more complex task which makes a definite appeal to the child. More specific still are such important questions as: (1) What is the best order of teaching the different stitches used in hand sewing? (2) Does the early introduction of machine sewing interfere with the formation of right habits of hand work? (3) At what age is it most economical to commence instruction in sewing? (4) What value has previous home teaching as an introduction to school work? (5) How should wrong habits, formed by such teaching, be broken by the school?

Thorndike's four rules of common sense occur to one as suggesting many rather simple experiments which could be performed to test controversies concerning methods of teaching.

His first rule, "Form habits, do not expect them to create themselves," suggests that we determine whether or not it is better: (1)

to push the needle with the side rather than with the end of the thimble, (2) to move the material up and down with the left hand, holding the right still in gathering, (3) to hold the thread taut between the third and fourth fingers of the right hand rather than with the fifth finger. These and many other points upon which sewing teachers disagree suggest that we determine the usefulness of the habit formed in one way or the other, and henceforth teach in that way.

Rule 2, "Beware of forming a habit which must be broken later," suggests balancing up present against later advantages. Consider, for example, such practices as (1) having the first work in sewing done with double thread, (2) drawing a line for the child to follow in her early attempt at running and backstitching, (3) having the work at first always started by the teacher.

Rule 3, "Do not form two or more habits when one will do as well," suggests interesting experiments to discover whether one will "do as well" in the following and other cases: (1) Taking several stitches upon the needle at one time in running; (2) measuring and breaking the thread with one movement; (3) making knots with one hand only; (4) taking several stitches upon the needle at one time in overcasting.

Rule 4, "Other things being equal, have a habit formed in the way in which it is to be used" again suggests testing out certain methods which would apply this rule, and sometimes finding out when the "other things are equal." We find that children often hem with the material held upside down. (1) Is hemming likely to be done in this way if it is used early upon a large garment, where the awkwardness of this manner of sewing shows itself? (2) Is damask hemming more likely to be put upon other kinds of material when it is taught only upon damask than when it is first used upon other materials? (3) When overcasting is first put upon articles whose edges are liable to the strains of actual wearing and laundering, is it so likely to be made in such a manner that the stitches are not deep enough for practical purposes?

The brief suggestions made here in accordance with the four rules of habit formation are for the most part, as the reader sees, applicable to the evaluation of habits which would appear in simple stitch making. If the reader will take the trouble to apply them to the more complex situations in which drafting, cutting, and fitting together are concerned, it will be seen how suggestive these rules

are, and how useful it would be to our everyday practice if many of them were actually tested out as they apply to the concrete sewing situations.

Besides these problems of method and procedure in sewing instruction, knowledge of norms of sewing ability is needed. We need to know how well children of different ages, school grades, and with various amounts of sewing instructions *do* sew. We need to decide also how well they *should* sew. This latter problem is one to be solved partly by others than educational psychologists. To some extent the answer, however, should depend upon a knowledge of the sewing ability of otherwise efficient and useful members of the adult woman population of which the school girls will one day form a part. If this ability is in general lower than that required by most of our schools for girls in the grades and high schools, it would be well, perhaps, to lower or revise our sewing requirements. This is precisely what Thorndike found in his study of merit in handwriting, and he wisely suggests that as adult members of society are able to succeed with a smaller amount of merit in handwriting than that required of the school population, the latter requirement is probably excessive and the time spent by school children in perfecting their ability along this line would far better be spent in learning to typewrite, an art which is quite rapidly replacing need for ability, beyond a certain point, in handwriting. Machine sewing is, in the same way, rapidly making unnecessary much of the ability for very fine work in hand sewing; and if school children are being required to do much better work in hand sewing than the normal performance for competent women, it is time that we realize the fact. Knowledge of this point can come about only through scientific measurement of sewing ability.

These much-needed norms of the typical amount of sewing ability characteristic of various ages, school grades, and sewing trades, expressed in terms of the scale presented here or of some other scale, must, of course, in some way take account of one other variable, namely "speed." How *quickly* a girl sews obviously is often as important as, or more important than, how *well* she sews. The measurement of speed will not be discussed in this monograph, not because the matter is unimportant, but because the procedures are obvious.

The questions of method and procedure which have been enumerated above have been, and still are, subject to debate among various

members of the household arts profession. Deductions (probably often very well drawn) from other fields have been introduced as proofs of various beliefs, but no actual experimental evidence is furnished (so far as the author knows). Indeed, no actual experimental work which might prove or disprove any of these theories is likely to be made until some means of measuring the results of instruction in sewing are at hand. This present study aims, among other things, to furnish the means of measuring some of the results of sewing instruction. The word "some" is used because it is recognized, first of all, that, as in other subjects, instruction in sewing aims at many other goals besides that which its specific title implies. Such aims as ideals of workmanship and the love of truth and honor, this research in no way helps one to measure. A section is included in the Appendix which may throw some light on the rather general aims of teaching "neatness" and "æsthetic appreciation," but, on the whole, the measurement here attempted is the measurement only of the specific aims of instruction in sewing. Even these are by no means all provided for. In fact, only a small part of elementary sewing is measured. This study still leaves us in need of instruments which will serve as measuring rods for the important abilities of cutting out, fitting together, planning and constructing various articles, as well as those which will measure the sewing on of buttons, the making of buttonholes and other finishings, and stitches which are not included in this study.

The author fully realizes that the contribution here made to the measurement of school sewing is small, indeed, when compared to the possibilities. However, when we consider, not the extent of the whole subject of school sewing, but the prevalence of the teaching and the need for the teaching of that part of elementary sewing for which a measure is provided by this research; when we take into account that most women do mending for themselves or others, that many women make some of their own or children's garments, and that approximately one out of every five or six girls[2] in our big cities enters some branch of the sewing trades, it seems to be desirable that even a meagre beginning be made in the scientific measurement of the school preparation for these life activities.

[2] This proportion was found to be true in Cleveland by the Survey Committee of the Cleveland Foundation in 1915.

The specific tasks of this research are as follows:

1. To make a scale by means of which merit in certain forms of hand sewing may be measured.
2. To make an inventory and analysis of the faults found in children's sewing.
3. To determine the relative importance of various faults and of various stitches as indicators of general merit in sewing.
4. To determine the relative reliability of judgments concerning various faults and concerning various stitches.
5. To determine (*a*) the reliability of one or more samples of a child's sewing in evaluating her real ability in that kind of sewing; to determine (*b*) the number of samples of various stitches which would be required in order to have various degrees of probability that a child's real ability to execute that particular stitch was measured.
6. To determine the relative accuracy of competent persons, with varying degrees of knowledge about sewing and the teaching of sewing, as judges of sewing products.
7. To test the value of our sewing scale as a means of attaining greater exactitude in the grading of sewing products.

SECTION II

THE DATA

SUBJECTS

The sewing of 1,212 individuals forms the material for this investigation. Ninety-five of these subjects were women students of colleges of practical arts, eight being especially chosen because of their known ability in hand sewing; eleven were adult women, friends of the author, whom she supposed to be above the average in sewing ability; thirty-seven were feeble-minded or defective children (both boys and girls) who were in special classes for defectives or in an institution for the feeble-minded; the remaining 1,069 subjects were girls ranging in age from nine to eighteen, and in school grade from the sixth to the twelfth. They represent fifteen different schools in six school systems, all in or near New York City.

It will be seen that, in obtaining subjects, an attempt was made to select more than a chance number of extremely good and extremely poor sewers. This is true to some extent not only of the adults, college students, and defective subjects, but also of the 1,069 regular school children. The request was made, in fact, to those who supervised the making of the samplers that they should choose their very good and very poor sewers to take part, if there was to be any choice at all within the class. This request was not made in the case of the sixty-eight children out of the 1,069 who each made two sewing samplers instead of one. In this case, we may presumably expect to find a normal distribution of sewing ability. For the group, as a whole, we should not expect this to be true. The selection was purposely made with a view to obtaining a graded series of specimens of sewing varying from as nearly perfect to as poor as it was within the author's power to obtain. Because the selection was made in this way, it has been deemed unfeasible in this study to draw any conclusion as to the normal sewing ability for various ages, grades, amount of sewing instruction received, etc.

MATERIAL

Each individual who took part in the experiment was given two pieces of longcloth. All pieces were of the same quality, six by three inches in size. To each teacher who supervised the making of these samplers was sent a model for the arrangement of stitches and the following directions:

DIRECTIONS

Material is 6 x 3 inches for easy handling, but only 3 inches of sewing, which is put 1½ inches from each end, is required, except in the case of the original basting which holds the two pieces together, and extends all the way across.

1. Use red thread No. 60.
2. Baste the two pieces together all the way across and do not pull out the bastings.
3. Do three inches of backstitching to form the seam.
4. Overcast the rough edges of the seam together for the three inches. Open the two pieces of the material and press back the seam so that the right side of the backstitch shows. Turn a quarter of an inch hem on the six-inch edge farthest from the backstitch and
5. Baste for three inches. Do not remove bastings from finished work.
6. Hem this for the three inches.
7. Do three inches of running stitch on top piece.
8. Do three inches of combination stitch on bottom piece.
9. Have pupils write on slip of paper name, age, grade, and number of years or months they have had sewing instruction. Pin this paper *securely* to their samplers.

N.B. The inclosed sample is in no way a model except for the arrangement of the stitches. Such things as knots, fastening of thread, manner of making combination stitch, etc., can be left to the individual.

These directions were not followed in many cases. In about 150 samplers several stitches were omitted or a half-backstitch was used in making the seams, and it was necessary to discard these, as it was feared they would introduce a variable feature which the judges would find hard to deal with. In some cases the directions to use red thread and to place the sewing one and a half inches from each end of the material were disregarded. These samplers were not discarded from the original judgments, but when some samplers were chosen for final placement, these doubtful ones were all omitted, so that the final scale contains only samplers in which the directions were closely followed.

In the case, also, of the sixty-four samplers which were judged for faults, an effort was made (as far as was consistent with other principles of selection to be explained later) to include only those samplers in which the directions were accurately followed. In only a few cases was it necessary to do otherwise.

In determining the reliability of estimates of sewing ability, based upon a limited number of samples, the work from one school was used. In this school each child made two samplers, but from among these 136 samplers only those 100 were used in which the directions were fully followed in both of the samplers made by each child.

Many forms of sewing might have been chosen other than the one adopted. For many reasons a completed article, such as a small bag or doll's apron, would have made better material for this study. The reader may wonder why we should revert to the sewing sampler which, as a teaching device, has for some years been discarded in our better schools. The reason is partly one of expediency. It was difficult to obtain 1,280 specimens of sewing made according to certain directions. It would have been far more difficult, if not impossible, to have obtained as many uniform specimens if the requirement had been to make them according to more elaborate directions, or according to directions which would require the spending of a much longer period of time.

Another justification for the use of a sewing sampler rather than a completed article is that it is more easily handled and probably appreciably better as a measuring device than any single specialized product would be. As a *teaching* device too great simplicity and formality are to be avoided. As a *measuring* device, provided they are elements of real situations, they are to be welcomed. This is not to gainsay the fact that other measuring instruments are also needed which will measure more complex abilities than can be reached by the simple scale; but it is to say that that which is measured (in this case, the making of certain stitches) is so done in a more accurate, neat, and concise manner.

In this case, as has been said before, no attempt is made to measure many of the complex abilities included in the making of any real article. What we do hope to measure is ability in the making of certain stitches, in the turning of the hem, and in the making of a simple seam (of a designated sort). The care of the sampler as a

whole as to cleanliness, mussiness, and trimmed edges [1] may be partly, though not very accurately, measured. The same may be said of the arrangement of the relative positions of the stitches.[2] These things do not, by any means, constitute total merit in sewing. Very probably, and to an increasing degree, as machine sewing comes more and more to replace hand sewing, some or all of them may become of minor importance. They are, however, of enough importance at the present to take the time and attention of hundreds of supervisors, thousands of teachers and millions of children in our country, on many days out of each week in the school year. Consequently, they are undoubtedly of enough importance to merit the existence of a scale which measures them alone, provided it does so in a concise and easy manner.

For this particular arrangement and choice of stitches on the sampler the author's judgment was the criterion. They seemed to represent better than any other arrangement and choice which was thought of, important sewing elements which are usually taught near the beginning of a sewing course, and which could be made in a minimum of time.

JUDGES

Three hundred and forty-seven persons made judgments upon one or another feature of varying numbers of the sewing samplers. The majority of these judges were women, but about fifty men and one child (aged twelve) also acted as judges. With not more than five or six exceptions, these persons were college students or graduates. Somewhat more than a random sampling of sewing teachers was included among the judges. The reader may wonder that a much larger proportion of sewing teachers was not included. The reason is threefold: (1) because of the difficulty in finding more sewing teachers who would make the judgments; (2) because what little data we have on the subject indicate that the judgment of sewing teachers is no more reliable than that of other intelligent

[1] The samplers with which we dealt, of course, changed in these particulars during the process of judging them. However, as they all had the same amount of handling, they presumably kept their original relative positions in these respects.

[2] To a large extent, this last matter of arrangement is prescribed by the directions, but not entirely. The directions, for example, say nothing about the placement of the running and combination stitches, from top to bottom of the piece upon which they were put.

judges;[3] (3) because it is possible that sewing teachers as a class may have a certain bias in favor of certain forms of sewing which would not be in agreement with otherwise competent judges in general. This may be an unreal danger, but those who have devised scales for handwriting, composition, drawing, etc., have taken pains to use the judgment of persons not all of whom were specialists in these subjects. The resulting scales have been found very useful by those who are professionally engaged in the teaching of these subjects, and it seems safe to follow the practice in question.

[3] Part of this evidence is given in Section 5. Another slight bit of evidence is found in the fact that when the sum of the deviations from the average judgments for ten samples taken at random was compiled, in the case of all the judgments made upon 129 samplers and upon 100 each of five stitches separately, the judge who was shown to be the best by having the least number of deviations was not a sewing teacher, nor had she ever studied sewing.

SECTION III

THE SCALE

The history of the making of scales to measure educational products is one of rapid growth since 1910, when Thorndike issued his scale for measuring the quality of handwriting. It is not within the scope of this study to present either the history or the value of this movement. Several books have appeared within the last year dealing with educational measurements which are in large part a collection and discussion of such scales. To these and to the original scales which they describe the reader is referred for general knowledge of the subject.[1]

In the Appendix of this volume is a reproduction of the graded series of sewing samplers which form the scale under discussion here. The judgments to be described which form the basis of the scale-values were passed upon the original samplers. The reproductions given in the Appendix are, however, by many persons considered so comparable to the originals that it is thought that their use will give substantially the same results which would be obtained if the original samplers themselves could be used by everyone. It is, of course, to be regretted that cleanliness and smoothness of the material are not well measured by this scale. They would not be, however, could the original samplers themselves be used, these having all become somewhat soiled and mussed by the amount of handling they necessarily received during the process of judging them.

There are three views of each sampler, all showing the stitches in their original size. One is a full view of one side of the sampler, and shows the way the stitches are arranged upon the original. The seam is here turned so that the right side of the backstitch shows. A second view gives the opposite side of the sampler, showing the reverse side of the hemming, running, and combination stitches. Before these photographs were made, the samplers were folded in

[1] Monroe, De Voss, and Kelley, *Educational Tests and Measurements*, 1917; Chapman and Rush, *The Scientific Measurement of Classroom Products*, 1917; Starch, *Educational Measurements*, 1916.

such a way that the running and combination stitches were brought closer together, and the ends of the material beyond the sewing were cut away. The stitches themselves are all included and are exact reproductions of the original. The third view gives a full-length view of the reverse side of the seam.

The fifteen samplers which constitute the finished scale are identified by the letters A, B, C, etc. The final scale values for each sampler are as follows:

A	0	I	8.9
B	1.8	J	10.4
C	3.2	K	11.5
D	4.3	L	13.1
E	5.1	M	14.3
F	6.1	N	15.1
G	6.7	O	16.4
H	7.8		

The scale should be used in the following manner: Any sample of sewing which is to be judged should be compared with the sampler in the scale which is most like it in *general merit*. It should be compared with several adjacent samplers until the judge is sure of its position. If it is exactly the same in general merit as some sampler in the scale, it should be given the number which is the score for that sampler. The score for each sampler is printed below it. If it falls between two samplers of the scale, it should be given a score between their values proportionate to the amount of its difference from each of them in merit. For instance, if an article to be judged has the merit of sampler L plus one-fourth the difference between sampler L and sampler M, it should be given a score of 13.4.

The essentials of any valid scale according to Thorndike are as follows:

1. Objectivity.
2. Consistency.
3. Definiteness of the facts and their differences one from another.
4. Comparability with the facts to be measured.
5. Reference to a defined zero point.

In making the present scale, the author sought to approximate these five ideals.

1. "Objectivity" is attained by having reproducible photographs of specified sewing samplers stand for or represent certain identi-

fiable positions in the scale so that it is possible for all persons to know what is meant by a sampler of a certain value.

2. "Consistency" is obtained by having a series of photographic plates of samplers which contain various amounts of merit in the same stitches, upon the same material, and with very much the same arrangement.

3. "Definiteness of the facts and their differences, one from another": the scale consists of reproductions of samplers, each one of which has a certain value, the samplers being arranged in order, starting with that one which has the lowest value. These values are in terms of units of amount. Each unit, which we may call K, represents that amount of merit which exists between the samplers when one is recognized as being of more value than the other by 75 per cent of competent judges and of less value by the other 25 per cent of the judges. To meet this requirement of definiteness of the facts as it has been met, the author has depended upon the coöperation of 111 judges who spent in all about 130 hours evaluating sewing samplers.

That the ideal of perfect definiteness is not attained is seen by inspecting the P. E.'s for each assigned value of fourteen of the samplers of those twenty-two which formed the basis of the first assignment of values*. The average of these P. E.'s is .159. This is .01 of the total range covered by the scale from the poorest to the best sampler, an amount of unreliability which is large, indeed, when we compare it with the definiteness of the units of such scales as those for weight or length. But if we could compare this amount of average deviation of the placement of values for each unit with the deviations which would result and which always have resulted, save by some chance occurrence, when sewing is graded for merit by one individual without, or even with, the aid of any scale, we should find .01 of the distance of our scale to be so small by comparison that we would regard it as almost negligible. The author does not mean by this that she is completely satisfied with the results. As time goes on and more and more judgments are available, it is her aim to make more and more definite the value of each sampler.[2]

*See page 28.

[2] Any reader who will have the fifteen reproductions of the samplers which appear in the Appendix cut apart and arranged in an order of merit by competent persons who are ignorant of their proper order or values, and who make their ratings independently, and will send these judgments to the author, will hasten the fulfillment of this aim.

Since the samplers themselves have not absolutely definite values, neither have the differences between them. Again, however, the variability of the judgments of individuals who may use the scale is sure to be of so much greater magnitude, that the usefulness of the scale is very little, if at all, impaired thereby. A greater drawback in the case of the differences is that they are not equal. This makes a little more ability necessary to use the scale, makes it harder to judge samples in integral numbers, or half units, and prevents a certain pleasant rhythmical feeling of evenness. More samplers and more judges alone can remedy this deficiency.

4. "Comparability with the facts to be measured." This principle of scale-making guided the author, as stated above, in selecting for the scale material samplers rather than finished articles because they seem to have more in common with all the various articles of sewing which they may be used to measure. A hem, a simple seam, and samples of running, backstitching (sometimes called stitching stitch), overcasting, basting, and combination stitches were chosen for the same purpose, because of the frequency of their appearance in the products whose measurement would require a sewing scale.

"Comparability with the facts to be measured" suggests another issue in the technique of scale-making. Many criticisms have been cast upon former scales because in the eyes of the critics either they have not been comparable to the facts to be measured, or they have included too many diverse elements which might vary independently in merit. Thus many have said that a scale for the drawing of houses could not possibly be used to measure the merit of the drawing of a snow-ball fight, and that a scale for merit in handwriting which included such diverse elements as slant of line and differences in spacing was useless. Results have proved these critics to be wrong. Such scales have been used, and to advantage. Yet, doubtless, other scales of a more comparable and analytical nature might have been even more exact in measuring the particular facts desired by these persons. The practical problem of making many scales of the necessary degree of comparability to the many different facts to be measured is a very extensive one.

Actual accomplishment must always be a compromise between what is feasible and what is theoretically desirable. Nowhere is this more true than in the making and also in the using of scales. This latter point should especially be held in mind by the critic of the analytic type. Granted that the present scale would be improved

if it were augmented by five more scales for each of the different stitches, would the additional time required to judge an article of sewing by the use of five scales rather than one repay the busy teacher by giving her enough additional definiteness as to the merit of the article which she was judging? A partial answer to this question will appear in Section 5. Suffice it to say here that when allowance is made for the difference in time spent in judging one set of samplers as compared with that spent in judging five sets, it seems evident, according to results which are recorded later, that a sewing sampler containing several kinds of stitches, a seam, and a hem is *in toto* nearly enough comparable to another such sampler to make it expedient to use a measure for general merit in that form, rather than a combination of independent measures for each of the separate stitches. If the fact to be measured is not general merit, but merit in overcasting, then undoubtedly a scale containing only samples of overcasting would be of greater value. The author is now engaged in making scales for the five different stitches (basting is omitted) which are represented on her full sampler scale. She is making these, however, with the expectation that they will be used more often separately to measure the particular kind of stitch of which it is a sample, rather than in combination in evaluating a complete article for total merit.

5. "Reference to a defined zero point" has been attained for the present scale by the following means: Eighty of the 111 individuals upon the basis of whose judgments the scale was made, were asked to locate the zero point of merit; in other words, to tell which, if any sampler, represented zero values. Zero was arbitrarily defined as meaning the score to be given a sampler which could be recognized as an attempt to conform to the directions for the making of the sampler, but which had absolutely no merit as an example of the stitches indicated. In dealing with sampler A as a zero, the reader must keep in mind just what is here meant by the term. Thirty-eight judges of the eighty who were asked to locate zero said that sampler A represented just zero merit, defined as above. Twenty-one of the judges thought that sampler A was better than zero, and twenty-one that it was worse than zero. Accordingly, sampler A just represents zero, as defined above, according to the combined opinion of these eighty judges, and it has, therefore, been given that value. It is interesting to note that this sampler was made by an imbecile boy, untrained in sewing, who is particularly

deficient in motor control and who suffers from serious eye trouble which often causes dizziness and nausea.

The usefulness of a scale whose values do extend upward from a defined zero point is found whenever it is necessary to make quantitative comparisons between the values of any two products which have been evaluated in terms of the scale. No one knows how much better the child who received 90 per cent as a school mark in sewing can sew than the one who receives 80 per cent; but in a very real sense it can be said that the child whose sewing has been rated upon this scale as 12.2, has produced a product with twice as much merit in the elements measured by the scale as is possessed by a specimen rated 6.1. The sense is, with certain limitations, the same as that in which it is said that four inches are twice two inches, the meaning being in both cases that the larger of the two contains twice as many units above zero.

Two methods of making scales such as this are in use by educational psychologists. They are based respectively upon the method of "right and wrong cases," and that of "mean gradations." The first makes use of the assumption that differences which are equally often noticed are equal; the second assumes that distances which appear equal to the average or median of many judges are equal. Each of these methods has points in its favor. The former has been chosen for the making of the present scale because it need not take account of the end error which must be reckoned with in making scales by the use of the second method. All the data necessary for assigning values according to the second method are in the possession of the author, who will be glad to give them to anyone who cares to compute the values in the other way.[3]

The following steps were followed in the making of this scale:

1. Twelve hundred and eighty samplers were examined by the author and all those which were made in accordance with the directions (or with only slight alterations) were retained.

2. Of these samplers, 854 were graded by twelve judges; 300 of them (which were received late) by only six judges.

The directions given to the judges were as follows: The samplers were to be placed in piles in such a way that the successive piles

[3] The general logic and presuppositions of these methods are assumed here and in what follows.

should contain samplers varying from one another in general merit by equal amounts. In all there were to be ten piles above zero merit (zero being defined as stated earlier in this section). If, in

TABLE I. *Distribution of 854 samplers as to general merit according to two groups, A and B, of six judges each*

Amounts of merit are in terms of six times the average value assigned the samplers by each group

Amounts of Merit	Frequencies for Group A	Frequencies for Group B	Amounts of Merit	Frequencies for Group A	Frequencies for Group B
6	2	2	37	32	38
7	1	3	38	19	31
8	2	2	39	39	26
9	2	4	40	30	25
10	3	3	41	21	21
11	2	2	42	36	34
12	3	2	43	43	32
13	3	3	44	19	31
14	3	4	45	23	19
15	1	5	46	24	15
16	5	4	47	23	20
17	11	3	48	18	25
18	9	5	49	16	10
19	4	7	50	8	9
20	9	8	51	5	6
21	18	12	52	5	10
22	23	20	53	5	5
23	14	21	54	9	4
24	15	31	55	3	5
25	17	16	56	2	1
26	17	22	57	7	3
27	35	22	58	3	8
28	25	29	59	2	1
29	29	27	60	2	1
30	27	29	61	1	3
31	31	25	62	1	
32	32	24	63	4	1
33	26	33	64		
34	19	40	65		1
35	28	32	66	3	2
36	34	27	67	1	

σ of dist. of A group = 10.51.
σ of dist. of B group = 10.34.

their opinion, some samplers were of just zero merit, these were to be put in an eleventh pile. If any samplers were of less than zero merit, they were to be put in the twelfth pile.

When it was found that none of the 300 samplers which were judged by only six judges fell within two piles of the top or bottom of the scale, according to the average opinion of the six judges, it was decided to use these no longer, as it was felt that the steps toward the center of the scale, which they did represent, were already sufficiently taken care of by the 854 samplers which had been more adequately placed. No further reference, therefore, will be made to these 300 samplers.

The twelve judges were now grouped according to the order of their judging, alternate judges forming Groups A and B. Table I shows the distribution of the 854 samplers as to general merit, according to the average opinion of each one of these groups. As the Table stands, the values are expressed as six times the average position assigned the samplers by each group. This expression is used to avoid the labor of dividing the sum of the six judgments all through by six, and also to avoid the use of fractions or decimals. Table I also states the σ [4] of each distribution. The correlation between the average judgments of Groups A and B was derived from the following form of the Pearson method for measuring correlation:

$$r = \frac{n\,(\sigma\, x^2 + \sigma\, y^2) - \Sigma(x-y)^2}{2n\,\sigma\, x\,\sigma\, y}$$

in which $\Sigma(x—y)$ is the sum of the squares of the differences between each pair of corresponding measures; $\sigma\, x$ is the σ of the distribution of the first set of six judges; and $\sigma\, y$ the σ of the distribution of the other set of six judges. This correlation of .93 $\pm$ .003 serves as a measure of the reliability of the judgments.

3. The value of each sampler was determined according to the combined average of all twelve judges. On the basis of these average judgments, the number of samplers to be retained for further judgments was now reduced to 129. These 129 samplers formed a rectilinear distribution, being chosen as representatives of equal intervals all along the scale from best to poorest, and differing from one another by amounts of one-twelfth of the differences represented

[4] σ will be used throughout to mean the Mean Square Deviation of a variable fact.

by any two adjacent piles into which the samplers had originally been sorted. This selection could not be rigidly adhered to near the two extremes of the scale. Occasionally there were gaps not represented by any sampler. To make up partly for these omissions five extra samplers from nearby steps were included in the 129.

4. These 129 samplers were arranged in groups according to merit by twenty-eight additional judges, in the same manner as that in which the 854 samplers had been arranged by twelve judges.

5. On the basis of the combined judgment of twenty-four of these last judges and of the twelve original judges, twenty-two samplers were chosen, the differences between successive average values of which were (with one exception, an extra sampler having been introduced due to a clerical error) approximately equal, except that the differences between the average values of two successive samplers at either end of the scale were smaller than differences in the middle of the scale and that one extra extremely good sampler was included. The reason for the choice of samplers near the ends of the scale which had smaller differences in average value than had those near the center was that the average value of the former probably gives a false estimate of their real value. The reason for this is that (according to the directions) no judge was able to place any sampler higher than pile 1 or lower than pile 12 (not even so low, unless he deemed it of less than zero merit). Therefore, the judgments of the extremely good and extremely poor samplers have distribution curves which are largely skewed toward the low and high ends of the scale respectively and which would probably be smoothed into more nearly normal surfaces of frequency if different conditions of judging had made this possible. The extra sampler mentioned above was included because it was hoped that by having a large number of extremely good samplers, one, at least, would be found which would fall approximately upon an even step of the scale.

All these preliminary steps were gone through in order to assure some degree of likelihood that the final values to be assigned to the samplers which should constitute the finished scale would be at approximately equal intervals from one another. The values so far arrived at were not used at all in the final scale except in a manner to be indicated later.

6. The twenty-two samplers whose selection has been described were next judged by fifty new judges. The placement this time

was in twenty-two piles; in other words, the samplers were arranged in an order of merit. Forty of the fifty judges were asked to locate "zero" by indicating which, if any, sampler was of just zero merit; or, if none was, whether they considered zero as falling

TABLE II. *Per cent of fifty judges making 'better' judgments upon a set of twenty-two samplers*

Sampler 544 is, in direct comparison with sampler 977, judged to be 'better' by forty-two per cent of fifty judges

977	is similarly judged	'better'	than	652	by	88	per cent	of	50	judges
652	is similarly judged	'better'	than	976	by	76	per cent	of	50	judges
976	is similarly judged	'better'	than	900	by	56	per cent	of	50	judges
900	is similarly judged	'better'	than	381	by	64	per cent	of	50	judges
381	is similarly judged	'better'	than	322	by	68	per cent	of	50	judges
322	is similarly judged	'better'	than	653	by	68	per cent	of	50	judges
653	is similarly judged	'better'	than	709	by	66	per cent	of	50	judges
709	is similarly judged	'better'	than	321	by	66	per cent	of	50	judges
321	is similarly judged	'better'	than	497	by	78	per cent	of	50	judges
497	is similarly judged	'better'	than	671	by	58	per cent	of	50	judges
671	is similarly judged	'better'	than	511	by	80	per cent	of	50	judges
511	is similarly judged	'better'	than	473	by	60	per cent	of	50	judges
473	is similarly judged	'better'	than	678	by	72	per cent	of	50	judges
678	is similarly judged	'better'	than	418	by	56	per cent	of	50	judges
418	is similarly judged	'better'	than	448	by	56	per cent	of	50	judges
448	is similarly judged	'better'	than	440	by	82	per cent	of	50	judges
440	is similarly judged	'better'	than	489	by	60	per cent	of	50	judges
489	is similarly judged	'better'	than	528	by	80	per cent	of	50	judges
528	is similarly judged	'better'	than	532	by	84	per cent	of	50	judges
532	is similarly judged	'better'	than	530	by	84	per cent	of	50	judges
530	is similarly judged	'better'	than	0	by	36	per cent	of	50	judges

between two samplers or as falling below the worst sampler. The judgments of these forty and of the first forty judges were used together in the manner described earlier to determine the fact that "zero" is exactly the amount of merit possessed by sampler A. The arbitrary definition of "zero" for our purposes is also described earlier.

7. As stated above, it had been decided that the final assignment of values for this scale was to depend upon the assumption that

differences which are equally often noticed are equal. It was necessary, therefore, at this point to arrange the data in such a way that a comparison between samplers of successive degrees of merit could be made in terms of the proportion of judges who made "better" judgments as to the relative merit of the two samplers constituting each successive pair. By a "better" judgment is meant one in favor of a sampler which had more merit according to the average opinion of the first thirty-six judges. Table II gives this arrangement of the data for the fifty judges in reference to twenty-two samplers. The original identification numbers are used to designate the various samplers in this table and in those which follow.

8. Although the forty judges who had judged 129 or more samplers probably had not made direct comparison of each of the twenty-two samplers with the next one below it, in the order given in Table II, it was possible to use their judgments in the same way as had been done for the fifty judges, with the exception that some provision had to be made for the "equal" judgments. The fifty judges, having placed the twenty-two samplers in an order of merit, of course had made no "equal" judgments; but the forty judges, having placed the samplers in only from ten to twelve piles, had often, it was found, placed adjacent samplers of our list as given in Table II, in the same pile. Table III gives the data and our manner of dealing with the equal judgments. Two plausible means of handling these were thought of, and are indicated here. One was to divide the "equal" judgments equally between the "better" and "worse" judgments. The assumption here is that since A and B were put in the same pile by the judge in question, he was as likely to judge $A > B$ as $A < B$. Another assumption, however, might lead to the conclusion that if the individuals who made "equal" judgments had been *forced* to make a decision between each of our adjacent pairs, they would have made judgments of "better" and "worse" in the same proportion as those judgments actually were made by those of the forty judges who did *not* make "equal" judgments. Since each of these assumptions has a certain degree of plausibility, it was decided to make a compromise midway between the two, as shown in the last column of Table III, and to use the amounts given there as the correct measures of the percentage of forty judges who made "better" judgments in respect to the first of each successive pair of the numbered samplers.

9. Twenty-five more judgments were next made upon the twenty-two samplers, by having them again placed in an order of merit. Twenty-one of these new judgments were made by additional judges, four by individuals who were among the first forty judges. Table IV shows the percentage of these twenty-five who

TABLE III. *The percentage of 'better', 'worse' and 'equal' judgments made by forty judges upon a set of twenty-two samplers, and the manner of dividing the 'equal' judgments properly between 'better' and 'worse'*

The table should be read as follows (each sampler being compared with the one below it): Sampler 544 was judged 'better' than sampler 977 by 7.5 per cent of 40 judges. It was judged 'worse' than 977 by 7.5 per cent. It was judged 'equal' to it by 85 per cent. When these 'equals' were divided equally between the 'better' and 'worse' judgments, it was judged better by 50 per cent of judges, etc.

Samplers each compared to the one below it	Per cent of 'better' judgments	Per cent of 'worse judgments	Per cent of 'equal' judgments	Per cent of 'better' judgments found when the 'equal' judgments are divided equally between 'better' and 'worse'	Per cent of 'better' judgments found when the 'equal' judgments are divided proportionally between the 'better' and 'worse'	Per cent of 'better' judgments found when the 'equal' judgments are divided as a compromise between 'equally' and 'proportionally'
544	7.5	7.5	85.0	50.0	50.0	50.00
977	27.5	2.5	70.0	62.5	91.6	77.05
652	35.0	5.0	60.0	65.0	87.5	76.25
976	48.7	15.3	36.0	66.7	76.0	71.35
900	53.8	12.8	33.4	70.5	80.7	75.60
381	50.0	30.0	20.0	60.0	62.5	61.25
322	44.7	23.7	31.6	60.5	65.3	62.90
653	40.5	18.9	40.5	60.8	68.1	64.45
709	42.1	18.4	39.5	61.8	69.5	65.65
321	54.0	21.6	24.4	66.2	71.4	68.80
497	44.4	25.0	30.6	59.7	64.0	61.85
671	50.0	23.7	16.3	63.1	67.8	65.45
511	36.8	31.6	31.6	52.6	51.8	52.2
473	56.4	7.7	35.9	74.3	88.0	81.15
678	48.7	20.5	30.8	64.1	70.3	67.20
418	55.3	21.0	23.7	67.1	72.4	69.75
448	48.7	15.4	35.9	66.7	76.0	71.35
440	40.0	17.5	42.5	61.2	69.6	65.40
489	50.0	17.5	32.5	66.2	74.1	70.15
528	45.0	12.5	42.5	66.2	78.2	72.20
532	42.5		57.5	71.2	100.0	85.60
530	30.0	12.5	57.5	58.7	70.6	64.65

made "better" judgments in favor of the first of each successive pair of samplers.

10. It was now necessary, in order that the final scale appear as a succession of samplers varying from one to another in merit by defined units of amount, that the measures we had of relative position be transmuted into measures in units of amount by means of the percentage of agreement of our different judges in respect to

TABLE IV. *Per cent of twenty-five judges making 'better' judgments upon a set of twenty-two samplers*

Each sampler is compared with the one below it. A 'better' judgment means that the first of the pair was considered better

Samplers Compared	Per Cent of 'better' Judgments	Samplers Compared	Per Cent of 'better' Judgments
544	40	671	64
977	88	511	60
652	44	473	84
976	88	678	24
900	36	418	80
381	72	448	64
322	64	440	88
653	60	489	80
709	56	528	88
321	60	532	96
497	72	530	

the relative positions. Since it was desirable, for reasons to be mentioned later, to keep separate for the present the three sets of judgments which we had obtained, these transmutations into units of amount were made separately for each of the three groups of fifty, forty, and twenty-five judges.

The theory underlying the transmutation is that differences which are equally often noticed are equal; that differences which are less often noticed than others are less than they; and that differences which are more often noticed are greater. Thus we say, judging by the results obtained from the fifty judges, that the superiority of sampler 977 over 652, which was recognized by 88 per cent of them, is greater than the superiority of sampler 652 over sampler 976, which superiority was recognized by only 76 per cent of these judges. The exact amount of each of these differ-

ences has been worked out (on the assumption that the distribution of the variations of individual judges' opinions of a sampler is that of the normal probability surface) in terms of a unit, which we may call K, which equals the difference between any two members

TABLE V. *The differences of twenty-two samplers between each successive pair, according to different groups of judges, as determined from a knowledge of the per cent. of judges who judged each sampler as 'better' than the one whose identification number lies below it*

K = a difference in merit, in favor of the first of each successive pair, which is noted by seventy-five per cent of the judges concerned

Samplers Compared	K, as Determined by Fifty Judges	K, as Determined by Forty Judges	K, as Determined by Twenty-five Judges
544	—.299	.000	—.376
977	1.742	1.096	1.742
652	1.046	1.058	—.224
976	.224	.836	1.742
900	.532	1.028	—.532
381	.694	.424	.865
322	.694	.488	.532
653	.612	.550	.376
709	.612	.598	.224
321	1.143	.728	.376
497	.299	.447	.865
671	1.246	.589	.532
511	.376	.082	.376
473	.865	1.308	1.472
678	.224	.661	—1.046
418	.224	.767	.625
448	1.355	.836	.532
440	.376	.587	1.742
489	1.246	.784	1.246
528	1.472	.874	1.742
532	1.472	1.575	2.596
530			

of a series just great enough to be recognized as being in one direction by 75 per cent of judges, the other 25 per cent of whom place the difference in the opposite direction. Thus, again according to the results obtained from the fifty judges, sampler 652 is slightly more than 1 K better than sampler 976, being considered so by 76 per cent (or slightly more than 75 per cent) of the judges, the remaining 24 per cent of whom reverse this judgment. By the

table [5] this amount is found to be 1.046K. By the use of the same table and our knowledge of the percentage of judges who made "better" judgments, as between each successive pair, all of the differences between successive samplers have been found. They are given here in Table V for each set of fifty, forty, and twenty-five judges. In Table VI are given the amounts of value for each sampler in terms of K, as deduced respectively from the fifty, forty, and twenty-five judges. These values have been found in each case by adding up from sampler 530 at the bottom (this sampler having a value of "zero" as explained earlier) the increments of difference shown in Table V between each sampler and the one next above it in order.

11. It is desirable always to have two separate measures of any series of mental measurements, in order that the unreliability of these measures may be determined in terms of the variability found between the two measures. To reduce our results to two comparable measures, it was here decided to use the results of the fifty judges as one measure and to average the results of the forty and the twenty-five judges for the second measure, giving equal weight to each group within the second measure. The forty judges are thus given less weight individually. The reason for this is that their manner of making the judgments, which necessitated the handling of many more than twenty-two samplers and probably prevented the direct comparison of the successive samplers in the list of twenty-two, made for less reliability of their results individually than did the more direct method of judging used by the groups of fifty and twenty-five judges.

The fifth column in Table VI shows the value of each sampler found by taking the average or median between its value according to the group of forty and its value according to the group of twentyfive.

12. The next step was to obtain a single value for each of the twenty-two samplers which should represent all the judgments made upon it. Column 6 of Table VI gives the measures of value obtained from all judges, by averaging the values according to the group of fifty judges with the values according to the combined groups of forty and twenty-five judges.

13. A measure of the reliability of the values of column 6 was obtained and appears in column 7 of Table VI as a list of P. E.'s

[5] A convenient form of which can be found in Thorndike's *Mental and Social Measurements*, p. 228. In this table what is here called K, is called Δ/P.E.

applying respectively to the value of the sampler opposite to which it stands. These P. E.'s for each value define the limits above and below the value, within or beyond which there is an equal likelihood that the *real* value (or that which would be determined by an

TABLE VI. *The amounts of value for each of twenty-two samplers, according to different groups of judges, found by adding successive amounts of differences from zero up*

Samplers	Value according to fifty judges	Value according to forty judges	Value according to twenty-five judges	Value according to the average of the group of forty and the group of twenty-five judges	Value according to the average between the group of fifty, and the average of the groups of forty and of twenty-five	P. E. of final amalgamated measure of value (Given only for the fifteen samplers to be used later)
544	16.155	15.316	16.028	15.672	15.96	
977	16.454	15.316	16.404	15.860	16.16	.181
652	14.712	14.220	14.662	14.441	14.58	.082
976	13.666	13.162	14.886	14.024	13.84	.109
900	13.442	12.326	13.144	12.735	13.09	
381	12.910	11.298	13.676	12.487	12.70	.129
322	12.216	10.874	12.811	11.842	12.03	
653	11.522	10.386	12.279	11.332	11.43	.058
709	10.910	9.836	11.903	10.869	10.89	
321	10.298	9.238	11.679	10.458	10.38	.049
497	9.155	8.510	11.303	9.906	9.53	
671	8.856	8.063	10.438	9.250	9.05	.120
511	7.610	7.474	9.906	8.690	8.15	
473	7.234	7.392	9.530	8.461	7.85	.373
678	6.369	6.084	8.058	7.071	6.72	
418	6.145	5.423	9.104	7.263	6.70	.340
448	5.921	4.656	7.858	6.257	6.09	.102
440	4.566	3.820	7.326	5.573	5.07	.306
489	4.190	3.233	5.584	4.408	4.30	.066
528	2.944	2.449	4.338	3.393	3.17	.136
532	1.472	1.575	2.596	2.085	1.78	.186
530						

infinite number of the samplers) lies. This was obtained, in the case of each sampler, by using two estimates of its value: one, the estimate made by the group of fifty judges; the other, the estimate made by the combined groups of forty and twenty-five judges. The average deviation of these two measures from their joint averages

was found, and this A. D. of the distribution was divided by the square root of the number composing it (here 2) in order to find the measure of reliability of the average. This measure of reliability, being in terms of an A. D., was multiplied by six-sevenths to turn it into a P. E. While each P. E. is thus itself subject to considerable unreliability, having been obtained from only two measures, the average of all the P. E.'s is a very reliable measure of the limits above and below the value within which each value as given would have an equal chance of falling were it to be obtained from an infinite number of judgments. This average for fourteen samplers is .159.

14. The amalgamated measure of value for the twenty-two samplers given in column 6 of Table VI served as the basis for the final selection of samplers which should constitute the completed scale. Fifteen of the twenty-two samplers were chosen for this purpose. They were so selected that the values between each successive pair should be as nearly as possible equal.

15. All of the steps which were gone through in determining the values of the twenty-two samplers after the judgments had been made upon them, were now gone through again for the fifteen samplers. The original judgments were used as it was desirable to compare directly each successive pair of the retained series. Table VII gives the consolidated review of this work for the fifteen samplers. It corresponds to the information given about the twenty-two samplers in Tables II, III, IV, V, and VI.

16. It now remained to compute one final single measure of merit for each of the fifteen samplers which constitute the final scale. In doing this we used the results obtained for the fifteen samplers in which we have directly compared the judgments of "better" made in the case of each adjacent pair, and also the results obtained when intermediate samplers were placed between some of the final ones. The working up of the data took account of these intermediate samplers. It was first thought that in doing this it would be better to give more weight to the values determined when the judgments compiled for the fifteen samplers alone were used, since samplers of intermediate amounts of merit being used in the other case might have caused some unreliability; but when it was found that the average for the P. E.'s was greater when the more direct method was used, in which fifteen only were evaluated, it was decided to give equal weight in the two cases. It must at this point be re-

TABLE VII. *A consolidation of the facts concerning the differences between successive pairs of fifteen samplers*

This table gives the same facts for fifteen samplers which are given for twenty-two samplers in Tables II–VI. A 'better' judgment means one in favor of the first of a pair—each sampler being compared with the one below it.

K = difference in merit, in favor of the first of each successive pair, which is noted by 75 per cent of the judges concerned.

Samplers compared	Per cent of fifty judges who made 'better' judgments	Per cent of forty judges who made 'better' judgments	Per cent of forty judges who made 'worse' judgments	Per cent of forty judges who made 'equal' judgments	Per cent of 'better' judgments (among forty judges) when 'equal' judgments are divided equally between the 'better' and 'worse'	Per cent of 'better' judgments (among forty judges) when 'equal' judgments are divided proportionately between the 'better' and 'worse'	Per cent of 'better' judgments (among forty judges) when 'equal' judgments are divided as a compromise between 'equally' and 'proportionally'	Per cent of twenty-five judges making 'better' judgments	K, as determined by fifty judges	K, as determined by forty judges	K, as determined by twenty-five judges	Value according to fifty judges	Value according to forty judges	Value according to twenty-five judges	Value according to the average of the group of forty and the group of twenty-five judges	Value according to the average between the group of fifty and the averages of the groups of forty and of twenty-five	P. E. of final amalgamated measure of value.
.977	88	27.5	2.5	70.0	62.5	91.6	77.05	88	1.742	1.096	1.742	17.511	13.914	19.740	15.827	16.67	.096
.652	76	35.0	5.0	60.0	65.0	87.5	76.25	44	1.046	1.058	—.224	15.769	12.818	17.998	15.408	5.59	.110
.976	70	72.5	10.0	17.5	81.2	87.9	84.55	88	.778	1.507	1.742	14.723	11.760	18.222	14.991	14.86	.081
.381	86	63.1	13.1	23.8	75.0	80.3	77.65	100	1.601	1.126	3.818	13.945	10.253	16.480	13.240	13.59	.214
.653	84	55.3	13.1	31.6	71.0	75.3	73.15	76	1.472	.916	1.046	12.344	9.127	12.662	10.894	11.62	.441
.321	90	55.0	22.5	22.5	66.2	70.9	68.55	84	1.900	.717	1.472	10.872	8.211	11.616	9.913	10.39	.292
.671	84	56.4	20.5	23.1	67.9	72.8	70.35	76	1.472	.793	1.046	8.972	7.494	10.144	8.819	8.89	.047
.473	82	76.9	17.9	5.2	79.5	81.7	80.6	68	1.355	1.278	.694	7.500	6.701	9.798	8.249	7.87	.228
.418	56	55.3	21.0	23.7	67.1	72.4	69.75	80	.224	.767	1.246	6.145	5.423	9.104	7.263	6.70	.340
.448	82	48.7	15.4	35.9	66.7	76.0	71.35	64	1.355	.836	.532	5.921	4.656	7.858	6.257	6.09	.102
.440	60	40.0	17.5	42.5	61.2	69.6	65.4	88	.376	.587	1.742	4.566	3.820	7.326	5.573	5.07	.306
.489	80	50.0	17.5	32.5	66.2	74.1	70.15	80	1.246	.784	1.246	4.190	3.233	5.584	4.408	4.29	.066
.528	84	45.0	12.5	42.5	66.2	78.2	72.2	88	1.472	.874	1.742	2.944	2.449	4.338	3.393	3.17	.136
.532	84	42.5	0	57.5	71.2	100.0	85.6	96	1.472	1.575	2.596	1.472	1.575	2.596	2.085	1.78	.186
.530	36	30.0	12.5	57.5	58.7	70.6	64.65		—.532	.557							

marked that an equivocal issue arose in securing the difference between samplers 381 and 653 in the case of the twenty-five judges, when the direct method of comparing judgments between the two was used. All of these twenty-five individuals judged sampler 381 as "better" than 653. Theoretically, it is impossible to assign an amount in terms of K to a difference which is noted by 100 per cent of the judges. The issue was met by assigning that amount of K which corresponds to a "better" judgment made by 99.5 per cent of the judges. Since the reliability of the table from which these values of K were derived decreases as we reach the extreme of 100 per cent of "right" or "better" judgments, the fact of the inclusion of this value of K when determining the final values for the fifteen samplers by the direct method is one more reason why the values determined by this latter method were not given extra weighting in the final consolidation.

The decision reached was that the final values as arrived at by the "direct" and "indirect" methods should be given equal weights. An average between them constitutes the final scale values for each sampler. They are as follows:

Original Sampler No.	Final Identification Symbol	Final Scale Value
977	O	16.4
652	N	15.1
976	M	14.3
381	L	13.1
653	K	11.5
321	J	10.4
671	I	8.9
473	H	7.8
418	G	6.7
448	F	6.1
440	E	5.1
489	D	4.3
528	C	3.2
532	B	1.8
530	A	0

SECTION IV

FAULTS IN SEWING

THE ANALYSIS OF FAULTS

The literature of educational psychology is each year becoming more replete with books and articles concerned with faults or errors made by children in the various school subjects. Such work is valuable to the educator in several ways, and various authors have studied faults with different purposes in mind.

Studies of children's errors are frequently made to aid the administrator in evaluating the success of various teaching methods and the personnel of his teaching corps.[1]

Studies of faults are often used in determining specific habits which should be formed. Freeman,[2] Scott,[3] McNamara,[4] and Pintner and Gilliland [5] have, by their investigations of faults in pupils' work, made valuable contributions to the teaching of handwriting, arithmetic, shorthand, and reading. Hollingworth, in a recent study of special disability in spelling,[6] has, among other contributions given us an analysis of spelling errors. She uses this as a basis for the study of improved methods for the teaching of spelling.

The mode of a child's or an adult's response, as determined by the *kind* of error or falling off from correct performance, has for some time been used in determining the degree of mental development reached. Binet [7] so used the type of response in defining words and evaluating other situations, and Freeman [8] and others show how mental development is reflected in the *kind* of departure

[1] Examples of such are found in Rice, J. M., "The Futility of the Spelling Grind," *The Forum*, vol. xxiii, 1897.

Judd, C. H., *Measuring the Work of the Public Schools*, Cleveland; Cleveland Foundation Survey, 1915, p. 290.

[2] Freeman, E. N., *The Teaching of Handwriting*, 1914.

[3] Scott, W., "Errors in Arithmetic," *Journal of Experimental Pedagogy*, vol. iii, 1916.

[4] McNamara, E. L., *The Methods of Teaching Shorthand*, 1914.

[5] Pintner and Gilliland, "Oral and Silent Reading," *Journal of Educational Psychology*, vol. vii, 1916.

[6] Hollingworth, L. S., *Psychology of Special Disability in Spelling*, 1918.

[7] Binet and Simon, "La mesure de développement de intelligence chez les jeunes enfants," *Société libre Étude psychologie de l'Enfant*, vol. xi, 1911.

[8] Freeman, F. W., *The Psychology of the Common Branches*, 1915, pp. 46–50.

from correctness which exists in the drawing of children and of primitive peoples.

A study of errors in children's school work is sometimes used by the psychologist as a means of defining better their normal process of learning. Such a method of treatment was adopted by Thorndike [9] in demonstrating the habits necessary for correct silent reading.

Certain as yet unpublished studies of Thorndike have made use of a knowledge of children's errors as a corroboration of what might be expected from lack of sufficient exercise of many important bonds. This lack he finds to be frequent, due to a common failing of most of our elementary text-books to give proper attention to exercising necessary bonds sufficiently. We may expect a movement for the revision and re-writing of text-books to come about through the impetus from such work.

Studies of errors and mistakes furnish, as is shown by Bronner,[10] considerable guidance in the individual diagnosis of unusual cases. Bronner advocates a more extensive use of this method, and its adoption in the study of the so-called normal as well as abnormal individual child.

As a general method for use in the study of heredity and brain anatomy many physiologists have long used the careful study of errors and mistakes in school subjects, particularly those found in reading.

This section, which deals with a study of faults found in children's sewing, follows only a few of the lines of research indicated above. For one thing, it deals with faults of the finished *product*, not at all with those of the process of sewing, except as these, of course, indirectly result from the latter. Our main problems have been the following:

1. To make an analysis of all the faults which the sewing sampler, which forms the material of our study, might contain, and to show how this analysis is useful in furthering some of the important conditions of improvement in sewing.

2. To find the distribution of the amounts of each fault present in a set of sixty-four samplers. This is done for all the faults when grouped under twenty-three headings, and also separately for six items making up one of these twenty-three headings.

[9] Thorndike, E. L., "Reading as Reasoning: A Study of Mistakes in Paragraph Reading," *Journal of Educational Psychology*, June, 1917, vol. viii, No. 6.

[10] Bronner, A. F., *The Psychology of Special Abilities and Disabilities*, 1917.

3. To find the correlation which exists between the amount of each fault which a sampler possesses and its total merit as judged in terms of our scale. This also is done for all faults when grouped under twenty-three headings, and separately for six items making up one of these twenty-three headings.

4. To find the correlation between the average amount of all faults possessed by a sampler and its total value in terms of our scale.

5. To find the correlation between the number of faults possessed by a sampler and its total value in terms of our scale.

6. To find the correlation between the average amount of all faults possessed by a sampler and the number of faults it possesses.

7. To find the comparative reliability of judgments made upon the different faults.

8. To find the correlation which exists between two of these facts, namely, the comparative reliability of judgments concerning various faults, and the comparative agreement between the different faults with general merit. We have sought, that is, to find the correlation between two correlations, the first of which is between two sets of judgments concerning the amount of the fault present, and the second of which is between the amount of the fault present (as determined by the average of all the judges) and the amount of general merit of the total sampler.

The analysis of faults which might exist in the particular piece of sewing we are dealing with was made in the following manner:

Eleven of the twelve individuals who made the judgments upon the 854 samplers were asked to record the factors which influenced them in making judgments, and to number these factors in the order of their importance in so influencing them. Table VIII gives the results of these eleven records. Very little attempt has been made to classify these replies, it being desired rather that the reader see them in their original wording, in which some are expressed in positive, some in negative terms. When two influencing factors were said to have equal importance they are given the same number in the table.

Another source of assistance in making the analysis was obtained as follows:

Several classes which were studying methods of teaching sewing and cooking in the practical arts department of Teachers College

TABLE VIII. *Factors influencing the judgment of eleven individuals in judging eight hundred and fifty-four samplers. Numbers refer to the order of importance of each factor to each judge*

	JUDGES										
	A	B	C	D	E	F	G	H	I	J	K
STITCHES											
Quality of stitch	2										
Evenness of stitch	3				1			3	5		
Size of stitch	5			4							4
Slant of stitch	6										
Good arrangement—crookedness of stitches	7										
Large irregular stitches and crooked line of stitching			3								
Fine straight even line of stitches			1								
Uniformity of size of stitches		1									
Spacing of stitches				8							
Evenness of line of stitching				5							
Unevenness of stitches				6							
Uniformity of slant of stitches		2									
Straightness of lines					2						
Backstitch loose						12					
Overcasting stitch loose						14					
Hemming stitch loose						13					
Hemming stitch too deep on hem						4					
Hemming stitches too close–go backward						5					
Hemming stitches too far apart						11					
Bad arrangement of stitches						15					
Small size of stitches, running and hemming		3									
Fineness of stitching							1	5			
Even spacing of stitching							3				
Straightness of stitching							2	2	4		
Comparison of the stitch 'as is' and 'as should be'								1			
Tension of stitch							7				
Length of stitch								6			
Execution of stitches—proper method, hemming backward									3		
Length of stitch in backstitch and combination									1		
Evenness in size and slant of stitches											2
Correct position of stitches											11
Evenness of lines											1

TABLE VIII—*Continued*

	JUDGES										
	A	B	C	D	E	F	G	H	I	J	K
KNOTS AND FASTENINGS											
Poor knots	9										
Poor ending or beginning of row	10										
Large knots showing and ends of threads hanging			5								
Insecure fastening of stitches so that sampler rips apart			2								
Big knots				2							
Absence of large and poorly made knots		7									
Ugly knots and fastenings					3						
No fastenings					10						
Well-fastened knots						8					
Knots							6				
Bad knots									4		
Sloppy fastenings											3
Knots on right side											5
THREAD											
Thread pulls so tight that sampler is puckered			6		7						6
Ends of thread				3							
Double thread					19						
Thread drawn tight						7					6
SEAM											
Thick bunchy seam			7								
Size of seam				10							
Evenness of seam				9							
Very large seams						9					
Very small seams						16					
Even seams							5				
Straightness of seams							6				
Straightness of hem or seam								4			
Straight seam										1	
HEMS											
Evenness of hem				11							
Edges not turned in hem						6					
Hem smooth, not puckered or drawn up		4									
Hem not turned straight at end						18					
Even turning of hem							4				
Uneven hem											8
Slant of hemming										1	
Hem on wrong side										8	

TABLE VIII—*Continued*

	JUDGES										
	A	B	C	D	E	F	G	H	I	J	K
OVERCASTING											
Overcasting too deep and close						2					
Overcasting thread not parallel						8					
Slant of overcasting										1	
Depth of overcasting										1	
PIECING OF TWO PARTS											
Straightness with which two pieces were sewed together			8								
Uneven piecing of two parts											7
NEATNESS AND CLEANLINESS											
Cleanliness and neatness	4	8			3						
Soiled, wrinkled and generally mussy appearance of sampler			4								
General neat appearance				1							
Dirty						17					
Neatness of cloth				7							
Cleanliness				7						6	
Rumpled cloth											9
Soiled cloth											10
Neatness—frayed edges not trimmed before overcasting										4	
General neat appearance									1		
Cleanliness of material									2		
GENERAL											
General appearance	1										
Failure to follow directions	8										
Incomplete—evidently slow, sometimes good	11										
Tightness or puckering of work				7							
Finishing off					4						
Evidence of effort					5						
Uneven, crooked sewing						1					
Edges not trimmed						20					
Understanding of thing to be done									3		
Spacing of lines on sampler											12
Following instruction—width of hem										2	
Completion of work										5	
Orderly arrangement—spacing										7	
Basting—a straight line		5									

TABLE VIII—*Continued*

	JUDGES										
	A	B	C	D	E	F	G	H	I	J	K
GENERAL											
Running or backstitching at uniform distance from line of basting		6									
Neatness and cleanliness, *i. e.*, smoothness of cloth, no evidence of much ripping out, cleanness		8									
Aesthetic effect, of even spacing, and beginning and finishing rows of stitches at exactly same distance from ends of sampler		9									

were visited and the students were asked to take home and fill in the following blank:

Write name, major subject, and whether you have ever been a teacher of supervisor of sewing (state which).

What are the five worst faults in the finished product of children's sewing:

(*a*) In order of badness—beginning with the worst.

(*b*) In order of frequency—beginning with the fault which appears oftenest.

It was particularly emphasized that the answers were to refer to the finished *product*, not to the *process* of sewing, and that the answers should be as specific as possible. It was also explained that the finished product they were to bear in mind was some small, simple article, as a small bag or sampler like the one used in our experiments.

The above blank and explanation were also sent to a class in the extension department of the University of Iowa. Forty replies were received in all, thirty-five from individuals who were at that time, or had at some time been, teachers of sewing. Table IX gives the results of these replies. Again very little attempt at classification has been made, it being preferred to retain the original wording of the replies. When two quite distinct faults were included under one heading, however, the terms are divided in our table but given the same number.

With these two sources of guidance and her own experience to draw from the author made a tentative analysis of all the faults

TABLE IX. *Chart showing teachers' judgments concerning five worst faults and five most frequent faults in finished article of a child's sewing*

First line opposite name of fault gives position of seriousness of fault
Second line opposite name of fault gives position of frequency of fault

(Thus the fault "seams too close to edge of material" is judged by Individual 2 to be first in seriousness and fifth in frequency of occurrence)

Individuals, all of whom have been teachers of sewing, except 10, 19, 25, 34, and 36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
SEAMS																																									
Too close to edge of material		1																																						
		5																																						
Uneven (as in French or fell seams)			1			2		4			5		3				3	3						2															
			1			3		3			5		2				5	5						1															
Careless seams (inside edge fraying and showing through as in French seams)					5																																		
					5																																		
Seam lines faulty (due to irregular stitching of seams)														4																									
														4																									
Badly clipped seams, especially French seams, before turning																			5																				
																			4																				
Crooked seams																				5						1	4					3							
																				3						3	3					4							
Seams pulling out because not taken in deep enough																							4																
																							2																
Seams not straight																													5										
																													2										
Ends of seams not fastened																																1							
																																5							
Frayed, irregular and too large French seams																																					1		
																																					1		
Fell seams not smooth, right side not stitched directly on the edge																																							4
																																							3
French seams—raw edges extending on right side																												4											1
																												4											1

TABLE IX—*Continued*

Individuals, all of whom have been teachers of sewing, except 10, 19, 25, 34, and 36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
STITCHES																																									
Uneven stitches																												3												
																												3												
Not evenly placed					2																				2							4								
					3																				1							4								
Not of even length	3		2	1	3		2							1		3		1	3					5	5			4			5		4						
	3		2	3	2		2							3		1		3						1	4			5			3		2						
Crooked or irregular stitches				2					5		3	2	1		3					4			5									3							
									5		2	2	4		4					2			4									4							
Size of hand stitches																																5								
																																3								
Too long or large										3							2										3												
										2							4										5												
Hemming stitch wrong side up								5															3																
								5															3																
Too long and loose to be strong															4																								
															5																								
Machine or hand stitching not close enough to edge of hem														5																		4							
Poor, coarse machine stitching													2	5																		5							
													3																											
Poor stitching						1		3																															
						1		2																															
Stitching poorly done on account of irregular tension, blunt needle, or carelessness in guiding machine																																				3			
																																				3			
Careless stitching either hand or machine																																					3		
																																					3		

TABLE IX—*Continued*

Individuals, all of whom have been teachers of sewing, except 10, 19, 25, 34, and 36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
THREAD																																									
Fastening thread insecurely (at end of seams)		2 4					1 1					4 3			2 1		1 3			2 4				3 2		3 4							1 3						
Using too long a thread		5 2								4 5																														
Pulling thread too tightly				3																																			
Poorly fastened threads																								2 4					2 4										
Thread soiled																																				4 2				
HEMS																																							
Uneven hems						3 4	4 4											3 5					2 1																	
Hems of varying widths					3 2																																			
Carelessly measured hems								2 4			2 1																													
Irregular turning of hems													3 2																											
Lack of exactness in folding hems																2 3																								
Poor hemming																																								
Hems (1) not stitched on edge, (2) fullness taken in by large pleats forming angular hem line																																						2 4	
Hem stitched in without actually catching hem into place																			5																	4 4			

TABLE IX—*Continued*

Individuals, all of whom have been teachers of sewing, except 10, 19, 25, 34, and 36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
KNOTS																																									
Making poor knots on thread, too big or too small		3 3																																						
Use of knots				4 2																										2 1									
Knots on surface					1 1																																			
Knots, bastings showing																4 2																								
Do not conceal knots																	5 2																							
Big knots at the beginning																				3 5																				
Large knots																								3 5					1 3										
Knots too large and dirty																																			3 3					
BUTTONHOLES																																								
Poor buttonholes and sewing on of button poor							5 5																																	
Poor buttonholes															5 2				2							2 4														
Bad buttonholes																		4 2															2 4							
GATHERS																																								
Gathers pushed together in places					2 4																																			
Uneven gathers							3 3				5 5	5 5																											5 5	

TABLE IX—*Continued*

Individuals, all of whom have been teachers of sewing, except 10, 19, 25, 34, and 36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
FINISHINGS AND JOININGS																																								
Careless finishing																																						3		
																																					2		
Bad joining of finishing edges						4																																		
						2																																		
Poorly constructed joinings												3																												
												4																												
Bad finishes																								1																
																								2																
Bad corners or finishing																		2																3							
																		4																5							
Finishing hurriedly done																																					2				
																																					2				
Finish clumsy and inefficient																																								5
																																								3
BASTINGS																																								
Failure to remove bastings																										5														
																										5														
Pinning and basting poorly done before stitching attempted					2																						2												
					4																						4													
SOILED WORK																																								
Work soiled			3	5									5				4			1			1	4	1	3			3			3		5						
			4	4									1				1			1			5	3	3	2			1			2		1						
Soiled and mussy appearance (crumpled)					1	5			1					3				5	1																					
						5			1					1				1	1																						
Work is not kept as clean as it ought to be																5																								
																5																									
Soiled product																																			1					
																																			4					

TABLE IX—*Continued*

Individuals, all of whom have been teachers of sewing except 10, 19, 25, 34, and 36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
CARELESSNESS AND LACK OF NEATNESS																																								
Carelessness in beginning or ending work	4																																							
	4																																							
Lack of neatness										2																														
										4																														
Neatness of finish											1																													
											3																													
Carelessness in putting things together																1																								
																4																								
Care of materials																						2																		
																						5																		
General lack of neatness and care																										4														
																										1														
Untidiness																														1										
																														2										
No definite ideas as to cleanliness of hands, neatness of work																															5									
Unneat appearance of finished article																															5		2							
																																	2							
INABILITY OR INACCURACY IN FOLLOWING DIRECTIONS																																							
Inability to follow directions as in width of seams					4																																		
					3																																			
Inaccuracy in following definite directions														2																				1						
														2																				1						
Inaccurate measurements												1		2	1																3			2						
												1		2	3																3									
Measurements																												1												
																												1												

TABLE IX—*Continued*

Individuals, all of whom have been teachers of sewing, except 10, 19, 25, 34, and 36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Position of needle and thread																					1 2 3 4 5 1 2 3 4 5																				
Position of needle and thimble																						3 3																		
One side puckered because work not properly held																									4 5															
Faulty positions of hand, fingers, use of thimble, etc.				.																										4 4									
Inability to handle single tools with ease, which leads to an amateurish-looking finished product																																		3					
Improper care of materials as shown by finished product, *e. g.*, cleanliness																																						2 2		
CHOICE OF MATERIALS AND COMBINATIONS																																								
Lack of individuality								1 1																																
Inappropriate material or not suitable combinations											4 4																													
Poor color combinations													4 4																											
Material not suited to weak, small fingers						.																					5 1												
Harmony																												2 2											
Material and trimming poorly selected																																							
Displeasing appearance as result of choice of poor, inappropriate material and design																																						4 4	

TABLE IX—*Continued*

Individuals, all of whom have been teachers of sewing, except 10, 19, 25, 34, and 36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Spacing in putting materials together																												5													
																												5													
Inexact cutting with pattern																			2																						
Inability to follow straight line																			4																						
																			3																						
Measurements following directions									4																															
									3																															
Cutting on straight of material									2																															
									4																															
General construction faulty due to omitting preparatory steps			4																																				
			3																																				
Errors as result of poor or improper cutting																																					5		
																																					5		
Cutting out inaccurately																																							1
																																							5
POSITION AND USE OF IMPLEMENTS																																							
Awkward use of tools, needle, thimble, scissors	1																																						
	1																																						
Awkwardness in holding material	2																																						
	2																																						
Bad position in sitting	5																													3									
	5																													1									
Dislike for using a thimble		4																																					
		1																																					
Poor handling of garment and care of work									3																															
									2																															
Material held incorrectly										5																														
										1																														

TABLE IX—*Continued*

Individuals, all of whom have been teachers of sewing except 10, 19, 25, 34, and 36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
GENERAL																																									
Attention																						1																			
																						1																			
Regularity of attendance																						4																			
																						4																			
Ability to keep work up-to-date																						5																			
																						2																			
Satisfied with too crude results, standard too low																											1													
																											2													
Heavy touch																														2										
																														3										
Desire to hurry work																														4										
																														5										
Poorly trained auxiliary muscles																														5										
																														4										
Lack of pre-thought to see some future steps depend on present one																															1									
																															1									
Child's dependence upon teacher for judgment of his work																															2									
																															2				1					
Unevenness										1																									1					
										3																									4					
Inaccuracy, ability, lack of neatness																																			4					
Uncleanliness																																			5					
Inability of some children to learn certain stitches																																			5					
Soiled condition of work due to carelessness in wetting thread, sewing with dirty hands																																				5				
																																				5				
Inaccuracy in form, *e. g.*, finished hems, seams, etc.																																					1			
																																					1			

TABLE IX—*Continued*

Individuals, all of whom have been teachers of sewing except 10, 19, 25, 34, and 36	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Lack of evenness and regularity in work																																								2
Lack of security and firmness in work																																								1
Parts of garments do not retain shape of original pieces																																							3
Material held incorrectly																																							4
																																							4
																																							2
																																					1		

which might be found in the sewing sampler under consideration. When she had done this, she and four other women candidates for the Ph.D. degree in Education at Columbia University went through her analysis step by step, adding to and subtracting from it. When it was complete they carefully read through the material given here in Tables VIII and IX to see that everything was included in their analysis which had been mentioned by the forty students and the eleven original judges, and which really are faults which might appear in our finished sewing sampler. The aim in making the analysis was to give a complete and also a mutually exclusive list of the possible faults. When finished it was read to a college professor and a college instructor, both specialists in the teaching of sewing methods, and was approved by them as being complete.

The final analysis which appears in Table X is thus the work of five students of Education, based upon suggestions coming from forty-three teachers (eight of these belonging to the original group of twelve judges) and eight other women, and approved by two specialists from the field of the teaching of sewing. In the opinion of the five students of Education and the two sewing specialists this analysis contains a list of all the faults which could appear in a finished sampler made according to our directions.

The purpose for which this analysis was made was twofold. In the first place, it was hoped that as it stands it will be a direct aid in the teaching of sewing. In the second place, this analysis was used, as will be shown under the heading "The Measurement and Significance of Various Faults," in a later part of this section, as the basis for a study of the number of faults which actually do exist in children's sewing under present conditions.

In order to justify the claim that this analysis may be of direct assistance in helping the teacher to bring about improvement in a child's sewing, which, after all, is the main direct aim of sewing instruction, the reader is asked to pause a moment to consider just what is meant by improvement in sewing. If we think of sewing ability as being represented by a horizontal line the left-hand end of which represents zero amount of sewing ability and the right-hand end of which represents perfection in sewing, and if we let the intermediate portions of the line represent a continuously increasing amount of ability from zero to perfection, then improvement in

TABLE X. *A detailed analysis of the faults possible in the sampler described in Section 3.*

Note. The technical meaning of stitches and spaces is here disregarded. A stitch here means the thread toward one as the sampler is held on the side on which the seam appears. A space is the distance between two stitches

1. Running Stitch:
   - Stitch too large
   - Spaces too large
   - Stitches uneven in size
   - Spaces uneven in size
   - Disproportion between size of stitches and spaces
   - Line of stitching crooked
   - Stitches drawn too tight
   - Stitches drawn too loose

2. Back Stitch:
   - Stitches too large
   - Back stitch not taken back far enough
   - Back stitch taken back too far
   - Stitches uneven in size
   - Line of stitching crooked
   - Individual stitching crooked
   - Stitches drawn too tight
   - Stitches drawn too loose
   - Occasional omission of back stitch

3. Combination Stitch:
   - Stitches too large
   - Spaces too large
   - Stitches uneven in size
   - Spaces uneven in size
   - Back stitching at irregular intervals
   - Disproportion between size of stitches and spaces
   - Line of stitches crooked
   - Individual stitches crooked
   - Stitches drawn too tight
   - Stitches drawn too loose

4. Basting:
   - Stitches too large
   - Stitches too small
   - Type of basting unsuited to purpose
   - Stitches uneven in size
   - Spaces uneven in size
   - Line of basting crooked
   - Stitches drawn too tight
   - Stitches drawn too loose

5. Overcasting:
   - Stitches too deep
   - Stitches not deep enough
   - Stitches of uneven depth
   - Stitches too close together
   - Stitches too far apart
   - Stitches at uneven distance
   - Stitches of uneven slant
   - Stitches of wrong slant
   - Stitches drawn too tight
   - Stitches drawn too loose

6. Hemming Stitch:
   - Stitches too long
   - Stitches too short
   - Stitches uneven in size
   - Spaces uneven in size
   - Stitches too close together
   - Stitches too far apart
   - Stitches of uneven slant
   - Stitches of wrong slant
   - Stitches too tight
   - Stitches too loose
   - Running stitch used, instead of stitch which holds fold down

7. Hem:
   - Too deep for appearance of whole sampler
   - Too narrow for appearance of whole sampler
   - First turn of hem turned in too far
   - First turn of hem not turned in far enough
   - First turn of hem unevenly turned in
   - First turn of hem not turned in at all
   - First turn of hem turned in too many times
   - Basting too far from edge of hem
   - Basting too near edge of hem
   - Hem turned down on side opposite seam

8. Knots:
   - Knots too large
   - Knots with loops
   - Knots too loose
   - Failure of knot to hold
   - Knot with loose end
   - Knot not concealed when possible
   - Knots in place of other type of fastening
   - Knots on side opposite seam

9. Fastenings:
   - Fastenings insecure
   - Fastenings conspicuous
   - Absence of knot or fastening at beginning
   - Absence of knot or fastening at end
   - Long threads hanging at either end of work

10. Seam:
   - Too deep
   - Not deep enough
   - Edges of material unevenly joined
   - Edges of material not trimmed before overcasting
   - Edges of material unevenly trimmed before overcasting
   - Sewing not extending to extreme edge of sampler when there is an attempt to do so.
   - Edges of seam turned over under overcasting

General:
   - Threads of material pulled making uneven edges (possibly due to blunt needle)
   - Lack of beginning and finishing rows of stitches at exactly same distance from ends of sampler
   - Poor spacing of rows of stitches from top to bottom
   - Thread snarled
   - Thread broken
   - Double thread used
   - Partly broken thread used
   - Extra material caught in under sewing
   - Wrinkled material
   - Soiled material
   - Evidence of ripping
   - Edges not trimmed

sewing can be conveniently thought of as advance along this line from left to right. If, instead of thinking of sewing ability as of a total, we break up the concept into the many elements which compose it, we may again picture each of these elements as a horizontal line, advance along any one of which from left to right will add its certain increment to improvement in the more general function of sewing ability. To think of sewing ability as thus spread out over a horizontal line is useful in that it emphasizes the fact that improvement in this function is improvement in the *amount* of ability possessed. To think of sewing ability as being composed of many elements each one of which may be represented by a horizontal line,

and improvement in any one of which is represented by advance along its line from left to right, is useful in that it emphasizes the fact that improvement in sewing is the sum total of improvements in the amounts of the *various* elements which compose it.

In what, then, does improvement consist, which makes possible advance along these lines? Thorndike, in his *Psychology of Learning*,[11] thus analyzes the concept for us: "Improvement is the addition or subtraction of bonds, or the addition or subtraction of satisfyingness and annoyingness. When any function is improved, either some response is being put with or disjoined from, some situation; or some state of affairs is being made more satisfying or more annoying. The rise of the practice curve parallels the growth of a system of habits, attitudes, and interests." By "bond" Thorndike means a connection in the nervous system between stimulus and response, such that when a given state of affairs arises a certain uniform response will automatically come about. Improvement is thus the forming and breaking of such connections, or, to use other words, the forming and breaking of habits of a very definite sort, plus the forming or breaking of likes or dislikes toward definite aspects of the work at hand. Thus, to form the connection between the sight of two pieces of material unevenly held together to form a seam, and the muscular adjustments necessary to bring about their proper placement, is to form a "bond," or many "bonds," the formation of which brings about their increment of improvement to the general function of sewing. More specifically, they bring about improvement in that element of sewing which might be called "the holding of the material," or even more specifically, "the placing of two edges to form a seam." This same connection between the sight of the two pieces of material unevenly held together and the muscular contraction necessary to bring about their proper placement, involves also the breaking of "bonds." "Bonds," which formerly led from this visual perception to the muscular contractions involved in continuing to so hold the material, must be broken in order that improvement may take place. Their subtraction, in fact, *is* part of the improvement. As for "the addition and subtraction of satisfyingness and annoyingness," which Thorndike says are part of what we call "improvement," these, too, are illustrated in the case under discussion; for if the child has learned to like or to be satisfied by evenness of the placement of the two edges, and to be dissatis-

[11] Thorndike, E. L., *Educational Psychology*, vol. ii, p. 186.

fied or annoyed by unevenness, an actual improvement has taken place which, other things being equal, will surely show itself when the next similar situation arises.

Keeping in mind these ideas of improvement as advancement along a line, or lines, which represent various amounts of an ability, or of its subdivision, and of advancement as constituted by addition and subtraction of bonds, and addition and subtraction of satisfyingness and annoyingness, let us see what light is shed by psychology upon the factors and conditions of improvement, and whether our analysis of the faults found in sewing may in any way aid in contributing to these factors and conditions. Again to quote from Thorndike's *Psychology of Learning*,[12] we find that as a summary of the "educational conditions of improvement" Thorndike says: "Assuming the acceptance of a certain aim for the pupil's exercise of a certain function, the selection, arrangement, and presentation of subject matter, and the approval, criticism, and amendment of the pupil's responses, are means of getting the pupils (1) to try to form certain bonds rather than others, (2) to form them in a certain order, (3) to identify more easily [13] the bonds he is to try to form, (4) to be more satisfied at the right bond, and more unready to repeat the wrong bonds, (5) to be more satisfied by the general exercise of the function, and (6) to be more satisfied by general improvement in it."

It is in connection with (3) and (4), above, that our analysis of faults may be of use. However, all six of these indications of how improvement can be brought about deserve close study, and because of their great importance all of them will be touched upon here.

(1) "To get the pupil to try to form certain bonds rather than others." In certain matters connected with the teaching of sewing, specialists are themselves not always agreed as to which bonds the pupil should try to form. To illustrate this disagreement I shall quote from the first four authorities whom I consulted, a part of their directions for the making of the overcasting stitch. No two of them agreed as to both the depth and distance of the stitches, and there is disagreement also as to the direction in which the sewing should be done. Brietzcke and Rooper [14] say, "Overcasting is

[12] Thorndike, E. L., *Educational Psychology*, vol. ii, pp. 230–1.

[13] Thorndike includes a footnote at this point which reads: " 'More easily' means throughout more easily than he would have done if left to his own devices."

[14] Brietzcke and Rooper, *Plain Needle Work and Knitting*, 1885, p. 21.

worked from left to right, the needle is inserted about four threads below the edge, the thread is passed over the edge. Miss three to four threads toward the right, then take four horizontal threads on the needle."

Patton [15] says, "Begin at the right-hand side. . . . Make the stitches one-eighth of an inch down, and one-fourth of an inch apart."

Smith [16] says, "Overcasting is worked from left to right. . . . Run the needle parallel with the edge for one-half inch, and bring out about one-eighth of an inch down."

Woolman [17] in discussing the same stitch says, "It is made usually from right to left (some prefer to make it from left to right). The stitches are of equal size, the depth and distance apart depend on the character of the material."

Such controversies must be settled within the profession, it being highly probable, however, that the use of our sewing scale may assist specialists in arriving at an agreement. In other matters, however, concerning which the members of the household arts profession are substantially agreed among themselves as to the final outcome which they desire to see fulfilled in a girl's sewing, there yet are great differences in teaching methods, which result to the pupil in the formation of very different bonds. The older analytic method of teaching stitches upon samplers, before their need was felt, resulted in the formation of bonds between learning the name of a stitch and making that stitch, or between having a small piece of material at hand and putting the stitch somewhere upon it; but it did *not* form the bonds between the real situations in which hems and seams and gatherings are necessary, and the response of making the proper stitch here in its place. To sew on dolls' clothes rather than upon a sampler, if the same amount of exercise be given, should lead the pupil to form quite as well the bonds required in actual stitch-making. It also should lead the pupil to form such bonds as that between the situation "a raw edge of material at the bottom of the skirt which if left would ravel out" and the response, "Make a hem of such a sort and sewed with such a stitch that it will be an added adornment to the dress as a whole."

Another highly important set of bonds, whose formation often is neglected in the teaching of sewing, are those in connection with the

[15] Patton, Frances, *Home and School Sewing*, 1901; teachers' edition, pp. 41–2.
[16] Smith, A. K., *Needlework for Student Teachers*, 1899, p. 90.
[17] Woolman, M. S. *A Sewing Course*, 1911, p. 48.

formation of habits of speed. To sew efficiently means quite as often to sew *quickly* as to sew *well*. School authorities responsible for the work in sewing too often neglect this fact or, if they recognize its truth, they fail at least to provide for it by seeing that pupils form the bonds necessary for the attainment of speed. The author sincerely hopes that her contribution to the subject of measurement in sewing, being, as it is, the means of measurement of *quality* in sewing, will never be construed to imply that achievement in quality, at the expense of speed, is always a desideratum in her eyes. Her hope, on the contrary, is that now that an objective measurement of merit is at hand, which is to some extent comparable to the clock as a measure of time, proper adjustment between the two desiderata, quality and speed, may be determined in terms of one another. She is fully alive to the plea of the investigators of sewing instruction in the Cleveland Survey and that of many other up-to-date educators that speed is often by far the more neglected of these two; and it is her earnest desire that her contribution may in some way be put to use to give this important factor its rightful place in our schools. Too long have those responsible for sewing instruction dillydallied with the notion that the merit of the finished product alone was important, too long neglected to give instruction in many methods of saving cross-cuts, and, above all, too long have they often rewarded rather than made annoying the useless habits of over-careful and punctilious work upon articles which would far better have been made quickly than made well.

(2) "To form bonds in a certain order" has probably been the real cause for the teaching of sewing by the use of samplers and for the undue emphasis upon precision of workmanship at the expense of speed. Some persons evidently believe that by forming first the bonds which have to do with stitch-making and *later* those concerned with the use of these stitches, the whole process of learning to sew will be facilitated. Unfortunately the time often never arrives when these latter bonds are formed. In any case, the assumption referred to certainly is not proved; experiment alone can absolutely prove or disprove it. Until such is made, however, common sense and experience in other lines besides that of sewing both tend to emphasize the fact that it is economical of both time and labor to "form habits in the way in which they will later be used," and that too great an attempt artificially to provide for the order of formation of bonds is apt to inter-

fere with the whole process of learning, due to the detrimental effect which such interference often causes in (4), (5), and (6) of the list of conditions of learning which we are now studying. However, if care is taken that these other conditions—(4), (5), and (6)—be not interfered with, it certainly is in accord with sound educational theory to expect that certain orders in the formation of bonds are preferable to certain other orders, and the average sewing teacher should be given more definite knowledge on this point than she now is given. Unfortunately, knowledge on this point beyond that possessed by sagacious guesses is not at present possessed by any one. The early physiological development of gross bodily control, as compared to the later development of ability to make fine muscular adjustments, suggests, of course, that the age of the pupil should be an important determiner of the order of the formation of bonds. Experimental evidence should be gathered to determine the proper age at which any sewing instruction should first be given, and the nature of the bonds which should be formed at various ages. If a girl, due to inner growth alone and apart from all training, so changes from the age of eight to the age of twelve that she is able to thread a needle as well, and as readily is able to profit by sewing instruction as is the girl of the same age who for four years has been given preparatory work in knitting and card sewing, let us recognize the uselessness of the latter preparations. If, on the other hand, these preparations are found by experiment to be of benefit, let us precisely measure the kind and amount of such preliminary training and of its benefit, to future ability in sewing.

Many supervisors say that evenness of stitches should be taught early, whereas the bonds producing smallness of stitches should be of later formation. The results of our experiment, to be reported in a later part of this section and which are recorded in Table XIII, show that among our selected samplers the fault of large stitches is not so great as is the fault of unevenness in stitches. This fact may be due to teaching which emphasizes the opposite procedure, namely, the teaching of small stitches early; or it may be due to the rather unusual manner in which our samplers were selected. As far as it goes, it indicates that pupils who vary from very good to very bad in their sewing are able as a group to conquer the fault of "stitches and spaces too large" as early, or earlier, than they are able to conquer the fault "stitches and spaces uneven in size." All of this simply illustrates our contention that knowledge is still lacking

which would enable one to know precisely in what order bonds should be formed in the teaching of sewing. Undoubtedly some preferred order exists which, when discovered, will lead to improvement in our methods of teaching sewing.

(3) "To identify more easily the bonds which the pupil is to try to form." It was in order to fulfill this condition of improvement that our analysis of faults, as given in Table X, was made. In order to make improvement in a general function such as sewing, it is very decidedly beneficial that the definite elements which constitute the general function, and along which progress should be made, be separately identified by both teacher and pupil. Of course the very existence of a fault implies its correlative virtue, and for some reasons our analysis might better have been made in terms of virtues or excellencies in sewing. As a matter of fact, however, such usage would have been rather pedantic. Any one looking at a number of sewing samplers and asked to comment on their merit, is very likely to be struck by the presence of faults. Had our analysis been stated in terms of excellences it would have resolved itself verbally mainly into a list of the absence of certain specified faults. For convenience, therefore, this negative handling of the matter has been retained. We must remember, also, that the analysis as it stands is intended for the teacher.

Pedagogically, however, it is advised that improvement be thought of by the child, at least, and to a certain extent by the teacher, as a positive, progressive matter. We hope not that "lack of unevenness," but that "greater evenness" shall be stressed; not that "crooked lines" be avoided, but that "straight ones" be the goal. The teacher, in fact, will note improvement often through the elimination of faults. Her special aim for one particular lesson or series of lessons may be to reduce "unevenness of stitches" among the class as a whole, or to help Mary Smith not to make such "large knots," or Elsie Jones to avoid "puckering of material due to hemming stitches being drawn too tight." None of these negative aspects need, however, be emphasized to the class. As a whole, they may be guided to the making of even stitches, this fact being often subordinated in their minds to their more immediate purpose of producing a beautiful and useful apron. Mary may be taught that small knots are made in such a fashion, and Elsie that smooth material in her finished hem will come about by pulling the thread just hard enough to cause it to lie evenly upon the material.

Psychologically, improvement in sewing is by our analysis reduced to improvements in the amounts of ninety-nine contributing factors. Each one of these factors is expressed in terms of a fault, the overcoming of which constitutes improvement in that element. According to our earlier description, sewing ability *in toto*, as far as it is expressed in the making of our samplers, is here thought of as represented by ninety-nine horizontal lines, each one of which symbolizes a particular sewing ability. The left-hand end of these lines stand for absolute absence of each particular kind of ability, or for a certain maximum of the fault in question. The right-hand ends of the lines respectively symbolize perfection in that particular element. The intermediate portions of the lines from left to right represent a continuously increasing amount of the element, when considered as an excellence, or a continuously decreasing amount of the element, if it is considered as a fault.

Quality and quantity in sewing ability are thus both cared for: quality by the existence of ninety-nine separate items; quantity by the existence of lines to represent each item; the lines signifying that various amounts of each item may be present. This representation of improvement in sewing can be made, although the faults themselves as given in Table X manifestly express somewhat different relationships among themselves. Some faults typically vary from zero to some maximum amount. This is true of "Overcasting stitches at uneven distances." Perfect evenness is the desideratum. From this lack of any of this fault, there may be increasing degrees of any amount. In the case of some faults, it is an all or none relationship; a fault, that is, either is not present at all or is present in its maximum amount. This is true of "Double thread used," "Hem turned on side opposite seam," "Running stitch used, instead of stitch which holds fold down," and some others. In such cases as these, our symbolic line must be thought of as consisting of only two points. Such cases, however, are rare, the possibility of intermediate amounts of each fault being the rule rather than the exception. What is worthy of notice is that in perhaps the majority of cases two antithetical faults exist, perfection being the golden mean between the extremes. Thus, in the case of "Overcasting stitches too deep," and "Overcasting stitches not deep enough;" of "Overcasting stitches too close together," and "Overcasting stitches too far apart." This possibility of erring in either direction from the desired goal is an additional reason why the analysis has been given

in terms of faults rather than in terms of excellences. "Overcasting stitches, correct as to their distance apart," is an excellence in sewing, but one which fails to call attention to the two ways in which failure to attain it often occurs.

In naming some of the faults, possibly a further analysis should have been made which would show this same possibility of deviating from perfection in two ways. "Line of stitching crooked," of course, refers to discrepancies which can take place either up or down from the golden mean of straightness. In some cases it is probably true that an individual child has a tendency always to sew "up" or "down," as the case may be. The teacher should be alive to this possibility, and in measuring faults which have this double implication she should herself, when necessary, amplify our analysis and direct the child's attention to the particular form of any fault which it is necessary that she be taught to overcome. It more generally will happen, however, that children have no individual bent toward "upness" or "downness" in their attempt to make a straight line, but that to a certain degree these two faults will both be present, making for a general "crookedness" in line. In such cases the pedagogical procedure should be to *avoid* the analytical suggestions.

The discussion of this last point may be taken as an illustration of a necessary comment concerning the value of the whole analysis of Table X. No analysis can be made which should invariably be followed in thought. There are times when more synthetic thinking is highly desirable. There are also times when further analysis should supplement the one at hand. There are many entries on Table X as it stands which should *not* generally be given piecemeal to the child. Even the teacher in many cases would be hindered rather than helped were she invariably to think of faults according to this very detailed scheme. Some faults are very often bound up with each other, so that the presence of one is quite likely to be accompanied by the presence of others. This is true of the three faults "Hemming stitches too long," "Hemming stitches of wrong slant," and "Hemming stitches too tight." The combination of these three faults produces the effect of an overhanding stitch having been made over the fold of the hem. Many teachers think of this as of *one* fault, and so deal with it in their teaching. This procedure is undoubtedly wiser in most cases than a more analytical treatment would be. On the other hand, it is quite conceivable, and often is actually the case, that one of these faults may be present

in some girl's sewing without the accompaniment of the other two. In such a case the teacher should be able to exactly state to herself wherein this child's trouble consists. To do so implies that she make or use an analysis. Good teaching must in the future, as it has in the past, depend largely upon a knowledge on the teacher's part of when an analysis of the process she is pursuing will help her pupil, and when it will not help. Good teaching will also depend, as it has in the past, upon the ability of the teacher to guide the pupil to the making of such analysis when it is felt that it would be an aid. A knowledge of general methods of teaching will help the sewing teacher in these two essentials. She also will be helped in both of them, if in her own mind the processes can be thought of both synthetically and analytically.

That teachers of sewing often do not think analytically of the faults of children's sewing, and that such an analysis as is given in Table X will really often aid in clarifying their own, and therefore their pupils' minds, as to the exact bond which the pupils should try to form, will be best appreciated by a comparison of Table IX, which is a compilation of the opinions about faults of supervisors and teachers of sewing, with our analysis given in Table X. To tell a child that her "stitching is careless," that she has "poorly constructed joinings," that the "finish is clumsy," that she shows "lack of neatness" or "carelessness in putting things together," that her first, second, third, fourth, and fifth faults, in the order of their importance, are the "position of needle and thread," that "lack of individuality" and "harmony" characterize her work, and that she "has the inability to learn certain stitches" would not, to any great extent, identify for her the bond which she was to form. In Table VIII are given the factors which eleven judges stated as influencing them in their judgments concerning the sewing samplers. One judge, a specialist in the teaching of sewing, mentions the "quality" of the stitch. If such an unanalyzed statement were made to a child, would it not be difficult for her to form the proper bond to bring about improvement in this matter, and would it not aid her in so doing if she were told that instead of having poor "quality" her stitches were pulled too tight, or were too loose, or were uneven, or whatever the fault might be?

(4) "To be more satisfied at the right bond, and more unready to repeat the wrong bonds." In mentioning this fourth condition of improvement, Thorndike recognizes the second of the important

constituents of improvement. Improvement, as we quoted from him above, is not only "an addition and subtraction of bonds"; it is also "an addition and subtraction of satisfyingness and annoyingness." To be satisfied at the right bond, however, necessitates primarily that one recognize the right bond. At least when the results, due to the exercise of such a bond, are definitely recognized so that the accruing satisfaction may be assigned to its proper source, improvement is thereby much facilitated. All, therefore, that was said under (3) about the identification of bonds has weight here also. This cannot alone be attended to, however, by securing even a perfect analysis. Conditions (4), (5), and (6) all involve this important element of adding and subtracting satisfyingness and annoyingness; (4) has just been quoted; (5) is "to be more satisfied by the general exercise of the function," and (6) "to be more satisfied by general improvement in it." To like to sew better than one used to, and to be more annoyed when sewing is interfered with or prevented, is thus one element of improvement in sewing (5). To want to improve one's sewing ability so that one takes real satisfaction in any improvement which does occur, and feels annoyed by lack of such improvement, is also itself an element of improvement in the general ability of sewing (6). Such desire and satisfaction we may hope will be largely stimulated if pupils themselves use our sewing scale, and watch and record their own improvement in terms of this objective measure. To feel satisfied when some special contributing bond is exercised, also means improvement to the general ability in question. This is what is meant by (4) above. For instance, for a child to recognize that her stitches in backstitch are being pulled now with a proper tension so that they no longer will be so loose that a pin could easily be inserted under them, as was formerly the case, and for her to feel satisfied by the performance of this action which will result in making her stitch look right, or, in other words, at thus exercising the right bond, is an instance of improvement due to the attaching of satisfaction to the right bond. The teacher may, in the first place, have brought this about through praise of the proper stitch, or the child herself may have discovered it by inspection of desirable scale or other models. Satisfaction at the exercise of right bonds and annoyingness at the exercise of wrong ones are indeed, no matter how they have been brought about, important elements in the learning process of any subject. The actual muscular adjustments necessary to bring about correct

performance come to have a pleasant feeling of "rightness" or of familiarity. For the accomplished sewer to push her needle in a new or awkward way *is*, in itself, annoying. For the child to attach such a feeling of annoyance to an awkward act is for her to improve her ability.

In this account of improvement in sewing and of the conditions favorable to improvement, we have, at times, strayed somewhat far afield from the analysis of sewing faults. Since, however, this analysis was primarily made as a partial fulfillment of some of these conditions of improvement in sewing, it seemed justifiable to discuss to some extent the concept of improvement itself, and several of the conditions of improvement, in order that the exact place and purpose of our analysis might be made clear.

THE MEASUREMENT AND SIGNIFICANCE OF VARIOUS FAULTS

In order to find the intercorrelations which exist between total merit and the amount of each fault present in a sampler, it was necessary to have a certain number of our sewing samplers evaluated in terms of the amount of each fault which they possessed, and in terms of their amount of general merit. The number of faults given in the analysis (See Table X) was so great, however, that it was deemed impracticable to determine separately the amount of each one of these which was present in each sampler. A partial synthesis was, therefore, made from the list given in Table X. This synthesis was made by the author with the help of three of her earlier assistants. The aim was again to include all possible faults, in a mutually exclusive list, but this time to make the analysis less detailed. Table XI shows this new less-detailed analysis. The twenty-three faults there mentioned were now to be used as the basis for evaluating certain of our sewing samplers.

Sixty-four samplers were chosen from among the 854 which were originally judged by twelve individuals. The sixty-four were so chosen that sixteen equal steps of merit were represented, each by four samplers, according to the average of the twelve judgments which had earlier been passed upon them. Neither extreme of the original distribution of 854 samplers was represented, the range of the sixty-four samplers being from two and three-fours to nine and one-half, in terms of the original ten, eleven, or twelve piles into which all the samplers were first sorted. These sixty-four samplers thus fairly well represent all the grades of work which would ordi-

TABLE XI. *A less detailed analysis of the faults which are possible in the sampler described in Section 3*

1. Stitches and spaces too large
2. Unevenness in size: Includes
    1. Stitches uneven in size
    2. Spaces uneven in size
    3. Disproportion between size of stitches and size of spaces
3. Line of sewing crooked or not parallel with edge of material
4. Stitches drawn too tight
5. Stitches not drawn tight enough
6. Slant of stitches incorrect
7. Stitches not parallel in slant with one another
8. Knots and fastenings large, badly made, and conspicuous
9. Knots and fastenings insecure or absent
10. Edges of seam unevenly joined
11. Seam or hem too large
12. Seam or hem too small
13. Basting at wrong distance from edge of hem or seam
14. Hem badly turned in: Includes
    a. Unevenness of first turning
    b. Absence of first turning
    c. First turning too large
    d. Second turning uneven
15. Stitches unsuited for purpose for which intended
16. Threads left hanging. Broken and soiled thread and double thread used
17. Bad arrangement of stitches on samples
18. Wrinkled material
19. Soiled material
20. Edges not trimmed
21. Extra material caught in under sewing
22. Evidence of ripping and threads of material pulled
23. Individual stitches in back stitching or in combination stitches are crooked

narily be found in any schoolroom, the extremes of our 854 samplers, as the reader remembers, having been produced mainly at one end by college students specializing in household arts, and at the other end by children in special classes for defectives, most of whom had had little or no training in sewing.

Each of these sixty-four samplers was judged in reference to the amount it possessed of each of the twenty-three faults given in Table XI. At least sixteen judgments made by as many individuals were passed upon each fault in each sampler. Most of this judging was done by women in college classes in psychology. Each student was given a sheet with the identification numbers of the samplers upon it and blank places in which were to be written the student's name, class, and the name and identification number of the fault upon which judgment was being made. The name of one fault was then given to each student with the direction that judgments were to be made upon this fault only. The names of other faults were read so that the students would see that other aspects were to be cared for by other judgments. Whether or not other elements besides the particular fault under consideration entered in to influence the judgments it is impossible to say. All who took part in the experiment seemed to be interested and anxious to follow the directions, and many reported that it was not at all difficult for them to abstract the element in question and to be influenced by that alone. If the presence of other features did cause some constant error, it is again impossible to say in which direction this error would lie. It is conceivable that a sampler lacking merit in other respects might be placed lower than it should be in one special fault, due to the suggestions caused by the presence of other faults. It is also conceivable that a sampler excellent in other respects might be marked lower than it should be in some faults, because of the contrast represented by that fault with other merits.

The specific directions for scoring for faults were as follows:

The worst score which could be given for any fault was 5; the best score (which meant a total absence of the fault) was 0. Intermediate scores of 1, 2, 3, and 4 were to be given according to the amount of the fault present.

Preparatory to passing judgments the classes were shown samplers of the worst total merit, to give them some idea of the greatest amount of the faults which they might expect to find. They were told, however, that the list of faults had been prepared to cover all the *possible* faults which might be present, and that it was quite possible that some of them were not actually present in any of these sixty-four samplers, or were present in only a slight degree. They did not, therefore, feel obliged to give a score of 5 or 4, or possibly even of 1 to any sampler when scoring for certain faults.

Before judging, certain of the faults were explained in more detail. These certain faults were said to have meanings as follows:

Fault 10. "Edges of seams unevenly joined" referred to the edges of the material to be put together in making the seam.

Faults 11 and 12. Seam or hem "too large" and "too small," respectively, meant too large or small in respect to the rest of the sampler and the strength of the material.

Fault 15. "Stitches unsuited to purpose for which intended" meant that the proper *kind* of stitch should be used for various purposes, hemming stitch upon the hem, etc., and the presence of this fault indicated that the wrong kind of stitch had been used.

Fault 17. "Bad arrangement of stitches on sampler" indicated that the esthetic effect of spacing was bad.

Fault 20. "Edges not trimmed" referred to the raw edges of the whole sampler.

Fault 21. "Extra material caught in under sewing" meant that some part of the material which should not be sewed was caught in with the stitches. (This is a fault which is found more often in machine than in hand sewing.)

Fault 22. "Evidence of ripping and threads of material pulled" referred to marks left on the material from stitches which had been ripped out and unevenness of the material due to having caught or pulled one of the threads, as is sometimes done through the use of a blunt needle.

Attention was called to the distinction between faults 3 and 23, one referring to the crookedness of the line of stitches, the other to the crookedness of the *individual* stitch in backstitching and combination stitch. In running stitch all unevenness was to be attributed to *line* of sewing being crooked, it being impossible in connection with this stitch to draw the distinction between crooked line and crooked stitch except in this arbitrary way.

Fault 6. "Slant of stitches incorrect" was said to refer to hemming and overcasting stitches in which the slant might be at the wrong angle or in the wrong direction (as is the case when one hems "backwards"). This fault was distinguished from Fault 7, which referred not to faultiness of the general slant but to differences in the slants of adjacent stitches.

Besides these directions the students before judging were given a chance to ask questions about the fault for which they were to judge, and any student who felt especially incompetent to judge for the fault assigned was allowed to change with some one who did feel competent to make such judgment. The sixty-four samplers were then rotated around the class, each being given a score by each student. When all were scored, the names of faults were interchanged, and the samplers again rotated, each student now judging the samplers for a different fault.

When from sixteen to twenty judgments had been passed upon each sampler for each fault, the average of all the sixteen to twenty judgments was found, and served as the score for each sampler in each fault. Table XII shows the amount of each fault possessed by each sampler, the average amount of the twenty-three faults possessed by each, the number of faults possessed by each (found by arbitrarily calling a fault present only when there is enough of it to have given it an average score of at least 1.0, in the opinion of the sixteen to twenty individuals who judged it) and the value of the sampler in terms of general merit. This latter was found by having six judges score each sampler in terms of the scale, described in Section 3, and using as a final score the average of the six judgments. In scoring for faults a *low* mark means *little* of that fault. In scoring for general merit, a *high* mark means *much* general merit. We expect, therefore, to find a negative correlation indicated between general merit and each individual fault.

From Table XII one is able to construct for each sampler a mental image of the amount of sewing ability which it represents. If we picture to ourselves, as suggested earlier in this section, improvement as being the sum total of advancement along a number of horizontal lines—each one of which symbolizes some element of a complex ability—we can in this case make, for each sampler, a snapshot of the degree of perfection which has already been attained in each of the twenty-three constituents of sewing ability. Sampler No. 336 has thus 1.1 units of the way to go from 5—0 (the extreme limits possible according to our scoring method) before total absence of Fault 1 will be attained. Fault 2 is more nearly conquered, .38 of a unit of improvement only being required before perfection in that line will be reached. In Fault 3, .25 of a unit only is required; in Fault 4, .26. Fault 5 is already completely overcome (its score being 0). Fault 6 shows .75 of a unit yet to be attained, and so on through the list. If we turn our attention to the other end of the table, where are placed the samplers originally judged to be low in general merit by the first twelve judges of all samplers, and examine sampler No. 443 in detail, we find that the amounts of all but one of the constituent faults are greater than they were for sampler No. 336. The exception is in Fault 11, "Seam or hem too large."

TABLE XII. *Scores of sixty-four samplers in twenty-three faults and for general merit*

Identification No.	Value of whole sampler according to original twelve judges	Fault 1	Fault 2	Fault 3	Fault 4	Fault 5	Fault 6	Fault 7	Fault 8	Fault 9	Fault 10	Fault 11	Fault 12	Fault 13	Fault 14	Fault 15	Fault 16	Fault 17	Fault 18	Fault 19	Fault 20	Fault 21	Fault 22	Fault 23	Average of all Faults	Number of Faults	General Merit as judged by the Scale
336	31	1.10	.38	.25	.26	.00	.75	1.10	.28	.33	.45	1.30	.35	1.10	1.25	.26	.11	.37	1.00	.70	1.20	.00	.06	.45	.57	7	12.7
374	32	.54	.56	.41	.77	.05	.69	1.10	.18	.12	.19	1.29	.18	.76	.59	.29	.00	.42	.37	.71	1.80	.06	.11	.59	.51	3	12.9
382	34	.26	.77	.15	.21	.00	.35	1.00	1.10	.82	.23	1.59	.12	.76	.94	.37	1.95	.31	.75	.45	1.70	.00	.21	.61	.63	5	12.85
329	36	.76	.25	.25	.00	.05	1.17	.71	.55	1.39	.21	3.04	.35	1.50	.65	.23	1.10	.75	.33	.50	1.00	.05	.00	.32	.65	6	12.5
302	38	.45	.64	.50	.26	.05	.50	.71	.28	2.11	1.10	.73	.15	.36	.55	.33	1.60	.60	1.00	.50	1.60	.00	.23	1.55	.68	6	12.2
377(744)	38	1.10	1.70	.61	.70	.60	.23	.68	1.80	.67	.68	.95	.18	.83	.72	.55	1.00	.89	.89	.26	2.80	.00	.26	1.05	.83	6	11.65
665	39	.52	.77	.52	.30	.24	.76	1.20	.72	1.05	.47	.83	.37	.17	.50	.31	.31	.33	1.00	.63	2.90	.00	.16	.94	.05	4	11.8
377(511)	40	1.30	1.30	.94	.41	.29	.47	.50	1.60	1.12	1.05	.78	.70	1.40	.95	.52	.72	.62	1.40	1.30	3.10	.30	.29	1.19	.96	10	12.15
352	44	.55	1.10	.52	.22	.52	.62	1.10	1.10	2.12	.87	1.00	.56	.88	.94	.58	.83	1.66	.62	.47	2.10	.12	1.06	.75	.88	8	12.6
668	44	1.10	1.10	1.40	.65	.22	1.40	1.60	.91	.55	1.21	.48	.30	.42	.63	.85	.29	.70	1.20	.42	1.30	.52	.21	1.25	.81	9	12.05
746	44	.91	.83	.90	.47	.38	.71	.68	1.60	.18	.23	.88	.13	1.40	1.61	.43	.84	.50	.78	.33	2.00	.12	.28	1.17	.75	5	11.45
747	44	1.20	1.10	.56	.20	.22	1.00	.88	1.10	.11	.40	.71	.20	1.30	.63	.55	.23	.35	.89	.55	.83	.05	.53	1.52	.65	6	12.2
590	49	.95	2.00	1.10	.36	.38	1.17	1.10	1.70	.81	.39	1.22	.25	.47	1.00	.50	1.80	.66	1.00	.83	1.50	.06	.45	1.67	.92	10	11.75
112	49	1.20	1.70	1.20	.22	.38	1.20	1.20	1.50	.73	.54	1.42	.15	.61	1.57	.56	.95	1.78	1.40	1.48	1.30	.05	.60	1.13	.99	13	10.2
581	49	.82	.56	.81	.05	.11	.67	.52	1.10	.67	.42	.50	.26	.42	.68	.30	.94	.50	.66	.42	.70	.00	.00	.89	.52	1	12.25
734	49	1.30	1.20	.50	.42	.52	.71	.77	2.10	1.52	.39	.55	.50	.41	.72	.43	1.50	.55	.94	1.39	1.80	.06	.16	1.41	.86	8	10.95
244	54	.76	1.50	.64	.54	.58	1.19	1.50	2.00	1.05	.90	.90	.36	.85	.83	.52	1.00	1.50	.94	.58	1.20	.15	.39	.95	.91	8	10.9
367	54	1.60	1.20	1.70	.36	.95	1.12	1.10	2.00	.41	.41	1.88	.06	.43	1.05	.50	2.60	1.83	.33	.16	1.20	.00	.35	1.62	.99	12	11.15
662	54	1.40	2.10	1.40	.10	.37	1.60	1.80	1.10	.28	1.00	.64	.65	.75	1.10	.50	.88	.75	.78	.75	1.70	.00	.35	2.09	.96	11	10.3
751	54	2.10	1.80	1.80	.52	.73	.78	1.40	1.70	1.22	1.50	1.33	.56	1.70	1.28	1.00	1.20	.68	1.30	1.05	2.10	.18	.33	1.70	1.21	16	9.25

44	59	1.40	1.40	.45	.21	.73	1.12	.94	1.20	.65	.42	2.11	.56	1.20	1.33	.50	1.10	1.47	1.10	.70	1.10	.00	.74	.44	.91	11	9.6
412	59	1.90	2.60	1.70	.45	.16	2.05	2.20	1.10	.78	.90	.41	.42	1.30	1.05	.95	.50	.89	1.00	1.00	1.30	.00	.59	2.43	1.11	12	9.35
416	59	.91	1.90	.89	1.10	.24	1.95	2.20	1.40	1.12	.71	1.23	.75	.87	1.12	.62	1.10	.44	1.90	1.22	2.50	.12	.12	1.17	1.11	13	11.1
741	59	1.50	2.00	1.00	.36	.28	1.45	1.60	3.60	1.29	1.00	.81	.25	1.00	1.06	.85	3.80	1.00	1.10	1.30	2.60	.14	.41	2.10	1.32	16	8.95
125	64	1.40	2.30	1.90	1.10	1.00	1.83	1.60	1.50	1.17	1.00	2.44	.06	1.10	1.95	1.12	.73	1.68	2.50	1.00	2.20	.11	.53	2.17	1.40	19	8.75
694	64	2.60	2.20	1.60	.90	.22	2.40	2.40	2.00	.29	1.00	.86	.57	.73	.94	.85	.41	1.28	1.90	.84	1.70	.05	.72	2.68	1.26	11	9.35
212	64	1.10	2.50	1.31	.95	.32	2.23	2.50	1.30	.84	2.05	.45	2.20	1.40	2.20	1.22	2.40	1.91	1.50	.95	1.60	.60	.65	2.52	1.50	16	8.75
494	64	1.70	2.30	1.80	.68	.24	1.50	2.00	1.60	.78	.94	1.95	.06	1.10	1.39	1.12	.84	1.26	1.70	1.21	1.50	.00	.63	1.67	1.21	15	8.7
17	69	1.30	2.30	1.20	.86	.23	2.12	2.00	1.40	.35	.52	1.16	.26	1.00	1.82	1.00	.95	1.63	1.80	1.40	2.40	.14	1.09	1.91	1.25	16	8.95
499	69	1.60	1.90	1.80	.78	.67	1.41	1.60	.94	1.65	1.59	1.23	.12	1.60	1.72	1.12	.95	.89	1.60	1.28	1.90	.12	.61	2.77	1.29	15	8.7
605	69	1.60	2.00	1.90	.94	.72	1.55	2.10	2.10	1.72	.63	2.45	.40	1.10	1.16	.75	3.60	2.61	1.80	1.17	1.70	.05	.50	1.52	1.48	16	9.6
841	69	1.90	2.50	1.50	1.20	.55	1.50	1.90	1.30	1.44	1.05	.47	.68	1.40	1.52	.62	2.10	2.31	1.60	.17	2.50	.06	.39	1.70	1.32	16	9.55
48	74	1.80	2.40	1.90	1.30	.32	2.44	2.50	2.50	1.04	1.75	2.07	.39	.95	2.18	.86	3.40	1.83	1.80	1.09	2.57	.09	.85	1.77	1.64	17	8.0
415	74	1.70	2.40	1.80	.78	.86	1.59	1.90	1.40	1.23	.82	.59	.62	1.40	1.23	.62	2.20	.88	1.90	1.05	1.31	.00	.23	2.19	1.24	13	8.8
646	74	2.80	3.00	2.70	.90	.82	2.85	2.90	2.10	1.17	2.45	1.91	.31	1.90	1.79	1.13	1.70	2.15	2.30	2.10	1.28	.14	1.23	3.40	1.87	18	7.3
692	74	2.00	2.20	2.00	.28	1.36	2.55	2.60	3.10	3.89	.72	2.28	.18	1.20	2.17	1.41	3.40	1.94	1.30	1.05	2.00	.11	.58	1.89	1.74	18	8.7
9	79	2.30	2.50	2.30	.85	1.48	1.44	1.90	2.40	1.35	.89	3.47	.13	2.10	2.06	1.34	2.20	2.61	1.81	1.69	2.69	.06	1.00	2.29	1.77	18	6.85
146	79	1.90	2.50	1.90	.38	1.10	2.04	2.30	2.50	1.44	1.35	2.12	.20	.90	2.71	1.13	1.70	1.91	1.30	1.19	1.65	.20	.95	2.27	1.54	18	8.7
204(505)	79	2.30	3.10	2.60	1.80	.36	2.21	2.50	1.90	2.22	1.61	1.83	.18	2.20	1.53	1.48	2.30	1.31	2.50	.95	1.05	.44	.89	2.88	1.74	18	7.25
644	79	2.60	2.80	2.00	.50	.50	2.75	2.70	1.70	1.52	1.30	2.43	.40	1.60	1.70	1.25	1.20	1.95	1.50	1.28	.84	.14	1.06	2.65	1.58	18	7.35
3(13)	84	2.30	2.40	2.10	.65	1.00	1.76	1.80	1.90	.39	.89	2.00	.31	.58	1.44	1.43	.84	1.61	2.42	3.90	2.28	.00	.26	2.83	1.52	15	6.75
114	84	2.10	2.50	2.10	.28	.63	2.31	2.50	2.90	.70	1.57	2.58	.05	.85	2.20	.85	2.00	1.86	1.20	1.16	1.68	.00	.30	1.62	1.47	15	8.0
427(611)	84	2.50	2.60	1.90	.57	1.19	1.95	1.80	3.10	1.94	1.52	3.12	.12	1.60	2.11	1.31	1.40	1.47	1.70	1.00	1.76	.23	.84	3.05	1.68	19	7.7
636	84	1.90	3.80	2.90	1.30	.59	2.85	3.10	2.40	2.47	2.05	1.76	.73	1.90	2.50	1.20	3.60	2.52	2.50	1.28	1.50	.29	.45	2.68	2.01	19	7.05
11	89	2.20	3.80	2.30	2.30	.64	3.05	3.10	2.30	.78	1.00	2.05	.18	.82	1.44	1.93	.85	2.44	2.16	.41	2.50	.23	1.12	2.70	1.75	16	6.65
138	89	2.00	2.40	3.80	.76	1.05	2.27	2.00	1.00	2.30	1.57	3.24	.47	2.10	2.55	1.20	1.60	3.32	3.20	2.14	1.40	.05	2.20	3.00	1.98	20	7.0
308	89	2.90	3.00	1.80	.57	1.19	1.41	1.90	4.40	3.35	1.11	2.33	.75	2.10	2.28	1.60	3.50	1.57	.53	1.15	2.95	.05	.58	2.44	1.88	18	6.4
639	89	2.90	3.70	2.30	2.20	.26	3.15	3.60	2.90	1.19	3.33	2.67	.33	1.40	2.21	1.70	2.20	3.05	3.20	1.79	1.52	1.90	1.29	2.63	2.23	21	6.35
33	94	3.10	4.00	3.10	.50	.76	3.54	3.30	3.60	1.04	1.18	3.21	.04	1.50	1.90	1.80	3.20	2.76	2.50	1.95	1.95	.04	1.71	3.43	2.17	19	6.15
242	94	2.40	3.10	3.10	1.20	1.43	2.70	2.20	2.20	1.72	1.83	2.28	.25	2.20	1.95	2.40	2.80	3.05	2.00	.85	2.39	.05	.84	2.83	1.99	19	6.65
447	94	2.10	4.10	3.30	1.90	.50	3.80	4.10	2.60	2.12	2.37	.62	2.00	2.00	3.37	2.10	1.80	3.25	2.50	2.61	2.22	1.10	1.00	2.79	2.35	21	5.95
452	94	2.60	3.90	2.50	2.40	2.05	4.05	4.00	4.60	3.70	3.28	.55	2.30	1.90	1.89	2.70	3.70	2.10	2.60	1.79	2.47	1.12	1.25	3.33	2.64	22	4.75

TABLE XII. *Scores of sixty-four samplers in twenty-three faults and for general merit*—(*Continued*)

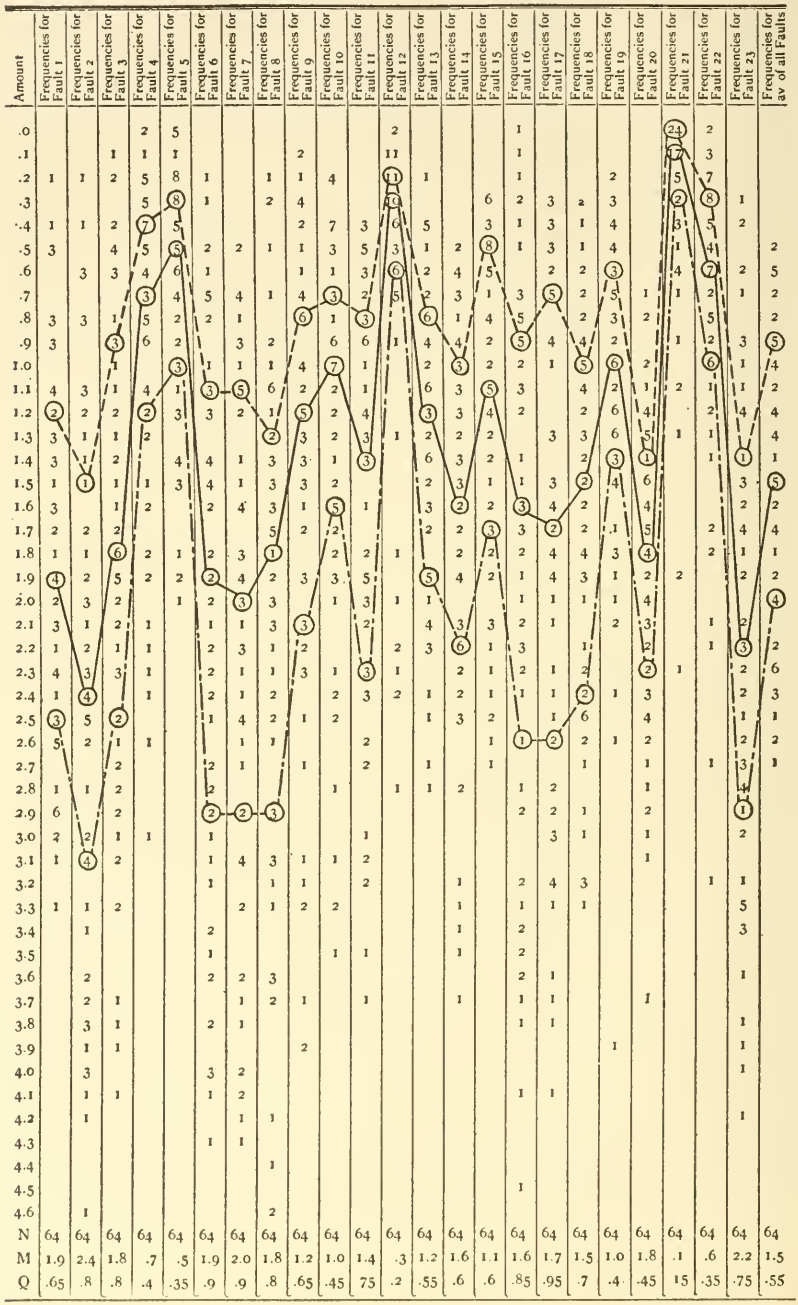
Identification No.	Value of whole sampler according to original twelve judges	Fault 1	Fault 2	Fault 3	Fault 4	Fault 5	Fault 6	Fault 7	Fault 8	Fault 9	Fault 10	Fault 11	Fault 12	Fault 13	Fault 14	Fault 15	Fault 16	Fault 17	Fault 18	Fault 19	Fault 20	Fault 21	Fault 22	Fault 23	Average of all Faults	Number of Faults	General Merit as judged by the Scale
444	98	2.30	4.00	3.90	1.90	.33	4.05	4.20	1.70	1.95	2.85	1.45	2.20	2.50	2.84	1.60	.72	2.85	3.30	1.50	1.53	1.35	1.76	3.33	2.35	21	6.0
59	99	3.00	3.10	1.80	1.10	1.95	3.16	3.60	2.90	2.72	1.95	.94	2.80	1.10	2.33	1.90	2.90	3.00	1.10	.29	1.50	.42	.76	2.18	2.02	19	5.0
247	99	2.90	3.30	3.00	3.00	.28	2.95	2.90	1.50	1.55	3.10	1.86	1.30	2.70	3.47	1.80	2.10	3.75	2.90	1.26	1.16	2.35	2.11	3.35	2.30	22	5.45
245	100	2.60	3.40	3.70	1.10	1.49	2.94	3.30	3.10	3.11	1.83	1.44	.50	1.70	2.11	2.10	2.30	4.11	2.70	2.05	3.69	.59	1.82	3.93	2.41	21	5.85
92	104	3.30	3.70	3.30	.75	1.54	3.84	3.80	3.70	1.33	1.64	2.70	.18	1.20	1.50	1.70	1.60	3.16	2.30	2.43	2.41	.00	2.72	3.77	2.28	20	6.45
442	104	2.90	3.10	4.10	2.10	1.19	3.64	3.10	1.70	2.21	1.88	.41	2.40	2.10	2.53	2.60	1.70	3.16	2.40	1.42	2.00	.88	1.41	3.44	2.27	21	5.6
613	104	3.00	3.80	2.20	.93	1.41	3.42	4.00	3.30	3.33	2.50	2.63	.33	1.90	3.68	2.20	3.30	3.77	1.10	.74	1.83	.58	1.75	3.28	2.39	19	5.15
634	104	2.90	4.00	2.90	1.80	1.86	4.16	4.30	4.20	3.89	1.67	3.11	.06	2.20	2.17	2.10	3.50	1.77	2.50	1.53	3.00	.59	2.00	3.61	2.60	21	5.5
137	109	2.50	3.60	2.80	1.60	1.82	3.61	3.10	3.70	2.28	1.89	3.66	2.40	1.10	3.22	2.50	2.90	2.88	3.20	1.47	2.12	.42	.82	3.25	2.47	21	4.8
258	109	2.90	3.60	2.70	1.50	1.38	3.41	3.70	4.60	3.23	1.41	1.95	.87	1.90	2.37	2.30	4.10	2.94	2.60	1.25	2.00	.16	1.00	4.16	2.43	21	5.6
275	109	2.50	4.20	2.50	1.60	.58	4.05	4.10	3.60	1.89	2.30	.91	.35	2.80	2.42	1.70	4.50	3.65	.65	.63	2.31	.70	.55	4.05	2.28	16	4.95
443	109	2.60	4.60	2.80	2.60	.33	4.37	4.60	3.20	2.29	3.59	1.09	1.80	2.40	3.32	2.50	3.20	3.21	3.00	1.80	1.58	1.90	3.16	3.32	2.74	22	4.85

The above comparison well illustrates two general principles worth considering, and suggests a third. One is the fact that on the whole correlation, rather than compensation, is the rule between the various constituents of total ability in sewing. Even when, as here, these constituents are mutually exclusive, it generally is the case that improvement in one of them is accompanied by improvement in the other, and that individuals possessing more merit in one element possess also more merit in the other. The cause of this correlation will be more fully discussed later. It is due to the presence of common elements which make for success in any constituent into which they enter. The importance of knowledge of this general correlation lies, for one thing, in the fact that we must recognize that to a certain extent good sewers are born and not made. Individual differences in capacity evidently exist here as in other fields. The wise teacher will recognize and allow for their existence. Teachers also should expect that in general, when improvement takes place, it should do so in *all* the elements which together make up sewing ability.

The second principle, illustrated by the scores of samplers 336 and 443, is also of importance. It is that positive correlation between constituent abilities, although the rule, is not by any means always 1.00. This is another way of saying that they are not illustrated in every case. This is shown above. Samplers 336 and 443, which were respectively judged as the best and as among the four worst of all the sixty-four samplers, when judged for general merit by the twelve original judges overlapped in their amounts of one element of sewing ability, sampler 336 having more of the fault "Seam or hem too large" than sampler 443. This fact, of occasional exception in individual cases to the general rule of correlation between the elements, brings up once again that which we sought to emphasize in the preceding section—that a careful analysis is necessary, and that each pupil's work individually should be thought of as made up of such and such amounts of ability along each of many contributing lines. Only through such analytic study will the diagnosis of the needs of each individual child be properly made.

A third principle, concerning the correlations between constituent elements of sewing ability, is that an equal amount of such correlation between each one of these constituents, as they are repre-

TABLE XIII. *Table of distribution of amounts of various faults found in 64 samples. The 25 percentiles, medians and 75 percentiles indicated by connecting lines. Data from Table X11*



N = Sums (M = Midpt.) $Q = \frac{75 \text{ percentile} - 25 \text{ percentile}}{2}$

sented by our list of twenty-three faults, and general merit in sewing, is by no means found. Inspection of Table XII suggests this. Just what the difference actually is will be shown in Table XIV, and its importance will be discussed in that connection.

Table XIII gives the frequencies for the amounts of each fault, arranged to show the medians and the 25 and 75 percentiles of each. These medians and quartiles are given at the bottom of the table. The facts brought out by this table should be studied by every teacher of elementary sewing. At a glance she may know what faults have the median frequency at some relatively large amount, thus indicating that they are overcome with difficulty and are likely to be present to a considerable amount in most children's sewing. Fault 2, "Unevenness in size of stitches and spaces and disproportion between size of stitches and spaces" is one such fault, evidently hard to overcome. Another is Fault 23, "Individual stitches in back-stitching and combination stitch are crooked." Thus by knowing the faults which are ordinarily present in the greatest amounts in children's sewing, teachers may prepare themselves in advance to combat these difficulties. By knowing the order of amounts of each fault which is ordinarily present in children's sewing, the teacher may know how to diagnose any individual irregularity. That sewing teachers do not already know these facts concerning the relative amounts of faults found in children's sewing may be seen by comparing Table IX, in which their opinions on this subject are given, with Table XIII, in which are given the obtained measures of the amounts of faults actually found in sixty-four samplers. These samplers were selected to represent a wide range in general merit, but in a random manner so far as comparative amounts of faults were concerned.

It was next desired to find the correlation which exists between each fault and general merit of the samplers.

The reader is warned here that when these correlations are found they will not in any instance represent the amount of *causative influence* which the presence or absence of any special fault has upon general merit. To believe such to be the case would be to ignore the fact that *all* sewing ability is probably positively correlated, and that one reason why general merit has a high correlation with the

absence of any fault is because both general merit in sewing and absence of that particular fault in sewing are each somewhat dependent upon other causes which influence them both. The method of Partial Coefficients has been designed to eliminate the effect of these extraneous influences. It has not been made use of here because the additional information it would give would seem hardly to justify the labor required to make the necessary computations. It is desirable, however, that the reader understand the method and the use to which it might be put, in order that he may not confuse the correlations which actually were found in this study of faults in sewing with those which *might* be found by the use of the method of Partial Coefficients. This method will be found worked out in Section V. The simple theory underlying it is well illustrated from another field. I shall quote here from a recent research [18] in which the problem is illustrated and the method of Partial Coefficients explained:

These results raise the important question of what the true relationship between algebraic ability and geometrical ability is, when these are freed from this common factor, verbal ability. How far is the correlation between algebraic and geometrical ability due to the correlation which exists between each of these and the abilities measured by the tests of what we have called verbal ability and to what extent is it independent of the latter?

To find an answer to this, recourse was had to the method of Partial Coefficients of correlation, by which the relationship between two functions for a constant value of a third can be determined.

The formula used was the following: [19]

$$r\ 12.3 = \frac{r\ 12 - r\ 13\ r\ 23}{\sqrt{(1 - r^2\ 1.3)\ (1 - r^2\ 2.3)}}$$

in which $r\ 12.3$ indicates the correlation between traits 1 and 2, for a constant value of trait 3. The reasoning underlying the Partial Correlation formula for three variables can be simply illustrated. Suppose that of the sixty-one Horace Mann students examined, ten are of approximately equal capacity in the verbal ability tests. The achievements in algebra and geometry of this group, in which verbal ability is constant, are then correlated. The resulting coefficient gives the partial correlation of algebraic and geometrical

[18] Rogers, A. L., *Tests of Mathematical Abilities and Their Prognostic Value*, 1918, pp. 81–82.

[19] Yule, G. Udny, *An Introduction to the Theory of Statistics*, Chap. 12.

abilities for a constant value of ability; that is, it expresses the relation of ability in algebra to ability in geometry independent of language ability; or, in other words, it represents the extent to which these abilities are related apart from their common connection with the ability to deal with words.

Dr. Rogers goes on to show how the method of Partial Coefficients may be applied to eliminate the effect of age as a second dependent variable. The method is theoretically applicable to the elimination of any number of extraneous factors. Its actual application, however, is extremely arduous when the number of factors to be eliminated is great. In our own case many such eliminations would need to be made before we could measure the exact effect upon general merit of any fault in any given amount. Here again ability to follow verbal directions and age are both common factors; so also are the abilities to stick to the task, to thread and manage the needle, and to control the handling of the material and the tension of the thread. To find articles produced under conditions in which these facts and countless other contributing features of ability in sewing were all constant, except for what we have called general merit in sewing and some one particular fault which we desired to correlate with general merit, would as the reader readily sees, require the collection of billions of sewing samplers. To make the elimination by the use of formulæ, without actually having the required samplers, would take much more time and sagacity in estimating the coöperating factors than the author has at her command. Again, then, we return to our original intention of finding the correlation which actually does exist between amount of each fault and general merit in sewing. We are aware that these correlations indicate no causative connection. They do, however, furnish important knowledge of how symptomatic any particular fault is of general merit, and for this reason a knowledge of these correlations will be found to be of great practical importance to teachers of sewing. With change of teaching methods it is quite conceivable that many of these correlations will change. With the present methods, however, as represented by the various schools in which these samplers were produced, these correlations hold, and when the twenty-three correlations between each fault and general merit are found, they may be used as indicating how significant any particular fault is as a symptom of general merit.

In order to find the real coefficient for this correlation of each fault with general merit, if the actual measures to be scored could be

freed from chance inaccuracies, we used the formula suggested by Spearman:[20]

$$r\,p\,q = \frac{\sqrt{(r\,p_1\,q_2)\,(r\,p_2\,q_1)}}{\sqrt{(r\,p_1\,p_2)\,(r\,q_1\,q_2)}}$$

in which $r\,p\,q$ equals the desired real coefficient of correlation, $r\,p_1\,q_2$ equals an actual coefficient of correlation found by pairing one chance measure of the first fact with one chance measure of the second fact, $r\,p_2\,q_1$ equals an actual coefficient of correlation found by pairing one other chance measure of the first fact with an additional chance measure of the second fact and $r\,p_1\,p_2$ and $r\,q_1\,q_2$ equal coefficients of correlation between the two pairs of measures for each of the two facts to be correlated. Table XIV shows the various correlations found. The faults are arranged in order according to the amount of the $p\,q$ correlation found. As we should expect, the correlation between amount of general merit and amount of each fault is negative; the amount of their importance, however, as symptoms of general merit, varies greatly. As we see, this variation is from an inverse correlation between Fault 2 and general merit of .94 to a similar correlation between Fault 20 and general merit of only .19.

This table makes prominent, to some extent, the same faults that Table XIII did, while both tables show some others as being very low in importance. Fault 2, "Unevenness in size of stitches and spaces," is seen to be most important as a symptom of general merit just as we found that the median amount of it which is present was greater than for any other fault. This fault is evidently the hardest fault to overcome, and at the same time the one whose elimination probably contributes much to general merit in sewing. So Fault 12, "Seam or hem too small," is given a very unimportant place in both

[20] Of course, as stated above, the selection of these sixty-four samplers was such as to bring about a rather unusual distribution of them when rated for general merit. The distributions of each fault, however, and of general merit also when this was estimated by six judges with the use of the scale (which measure was used in calculating our coefficients of correlation) were all approximately of the "normal type." The formula which we used for obtaining the correlations, $r = \frac{\Sigma\,x\,y}{\sqrt{\Sigma x^2}\sqrt{\Sigma y^2}}$, is therefore applicable. In any case, we used the same method for all the twenty-three faults, and our conclusions refer mainly to the *relative* differences between these.

tables, that is, it is shown by Table XIV to have little effect as a symptom of general merit inversely correlating with it only to the extent of .47, and by Table XIII it is shown to be present in only very small amounts.

TABLE XIV. *Coefficients of correlation found between general merit for the whole sampler and each of twenty-three faults*

p_1 = general merit as judged by three judges
p_2 = general merit as judged by three other judges
q_1 = amount of each fault as judged by eight judges
q_2 = amount of each fault as judged by eight other judges

No. of Fault	Name of fault to be correlated with general merit	rp_1p_2	rq_1q_2	rp_1q_2	rp_2q_1	rpq*
2	Unevenness in size	.967	.921	−.907	−.877	−.945
1	Stitches and spaces too large	.967	.775	−.851	−.774	−.937
6	Incorrect slant	.967	.909	−.822	−.896	−.915
15	Unsuited for purpose intended	.967	.795	−.705	−.903	− .91
3	Line crooked and not parallel with edge of material	.967	.811	−.832	−.756	−.895
7	Stitches not parallel in slant with one another	.967	.944	−.871	−.836	−.893
23	Individual stitch in backstitch and combination crooked	.967	.943	−.806	−.874	−.879
14	Hem badly turned in	.967	.892	−.761	−.818	−.849
17	Bad arrangement, etc.	.967	.917	−.750	−.803	−.824
18	Wrinkled material	.967	.836	−.747	−.618	−.755
16	Threads hanging, broken, etc., and soiled thread	.967	.760	−.601	−.693	−.752
10	Edges of seam unevenly joined	.967	.858	−.694	−.670	−.749
4	Stitches too tight	.967	.895	−.698	−.689	−.745
8	Knots and fastenings, large, etc.	.967	.904	−.698	−.695	−.745
5	Stitches not tight enough	.967	.818	−.663	−.633	−.728
22	Evidence of ripping and threads pulled	.967	.907	−.658	−.689	−.719
13	Basting at wrong distance from edge of hem, etc.	.967	.726	−.571	−.614	−.706
21	Extra material caught in sewing	.967	.629	−.557	−.325	−.545
19	Soiled material	.967	.792	−.438	−.518	−.544
12	Seam or hem too small	.967	.872	−.392	−.472	−.468
11	Seam or hem too large	.967	.709	−.439	−.273	−.418
9	Knots and fastenings insecure or absent	.967	.869	−.312	−.272	−.318
20	Edges not trimmed	.967	.784	−.286	−.101	−.195

* It is a recognized custom among statisticians in correcting for attenuation of minus correlations to disregard the exact working of the formula which would make all our correlations plus instead of minus.

On the other hand, Fault 15, "Stitches unsuited for purpose for which intended," is of rather great importance as a symptom of general merit (having an inverse correlation with it of .91), but it is not of very frequent occurrence, at least in large amounts, its median frequency for amount being only 1.1. Fault 4, "Stitches too tight," is another fault which happens very seldom—its median amount being only .7, but which still has a comparatively large amount of significance for general merit inversely correlating with that to an amount of .74.

Some other faults show an opposite tendency to these. Instead of occurring usually in only small amounts, but in spite of this, having a fairly large importance as symptoms of general merit, these faults to be mentioned occur in fairly large amounts, but have a comparatively small significance for general merit. Fault 23, "Individual stitches in backstitching and combination are crooked," is one of these. This fault has a median amount of 2.2, showing it to be greater in amount than any but one other fault; six other faults, however, have higher inverse correlations with general merit, Fault 23 having one of .89. Fault 11, "Seam or hem too large," Fault 9, "Knots of fastenings insecure or absent," and Fault 20, "Edges not trimmed," all are examples of faults which signify little as to general merit, having with it inverse correlation of only .42, .32, and .19 respectively, and yet their median amounts are about or greater than the average, being 1.4, 1.2, and 1.8 respectively. "Edges not trimmed," especially, is evidently not much of a fault, since in spite of its frequent appearance its negative correlation with general merit is only .19. Further inspection and comparison of Tables XIV and XV (until some better measures are at hand) should be made by every person engaged in the work of giving or supervising sewing instruction. Only by some such objective measure of the facts concerning sewing faults can the teacher know of their relative prominence and significance for general merit.

Table XV shows the coefficients of correlation between general merit, average amount of all faults, and number of faults, also the coefficient of correlation found between the correlations $q_1\, q_2$ and the correlations $p\, q$ as given in Table XIV. The high inverse correlation between general merit and the average of all faults—.94—indicates that we have probably included all salient features in our analysis, and that the relative importance of them is such that a

balance is fairly well secured when they are equally weighted. Only by the very elaborate method of "the regression equation" would it be possible to weight accurately every fault so that the correlation between average amount of faults and general merit would be an inverse correlation of 1.00. This, of course, would happen only if we actually included all faults, as we attempted to, and that in a mutually exclusive list.

TABLE XV. *Coefficients of correlation found between general merit, average amount of all faults, and number of faults, also between* $rq_1 q_2$ *and* rpq *of Table XIV*

Correlation between general merit and average of all faults,	$-.942 \pm .0095$ (P. E.)
Correlation between general merit and number of faults,	$-.924 \pm .0123$ (P. E.)
Correlation between average of all faults and number of faults,	$-.908 \pm .0148$ (P. E.)
Correlation between rq_1q_2 and rpq,	$-.891 \pm .029$ (P. E.)

The correlation of number of faults with general merit, and with average of all faults, has little meaning, since our definition of 1. as the average amount of a fault which is necessary in order that it be called present, is, of course, arbitrary. It was at first hoped that the number of faults actually found could be compared with the opinions of the thirty-six sewing teachers given in Table IX as to which fault is most frequent in children's sewing. Such a comparison cannot be made; we do not know at what average amount of a fault to say that it begins to exist. If we take the logical position, and say that it begins at immediately above zero, we shall have to say (judging from Table XIII) that all but six faults are present in all the sixty-four samplers. But our arbitrary decision to call a fault present when it has an average amount of 1. seems to be fairly justified by the fact that by the use of this measure the number of faults correlates inversely with general merit to the amount of .92. However, since this is no certain proof that our arbitrary decision has merit, the attempt to compare the most frequently occurring faults with those which in the opinion of sewing teachers were thought to be most frequent has been abandoned. Anyone caring to make this comparison may do so with the aid of the data in Tables IX and XII.

The correlations between q_1 and q_2 given opposite particular faults in Table XIV can be thought of as measures of the reliability of

judges in evaluating each of these faults. A high correlation between them for any fault indicates a comparatively large amount of agreement between judges as regards that special fault. The correlation of all these $r\ q_1\ q_2$'s with the $r\ p\ q$'s shows to what extent a high degree of agreement between judges as to the amount of any fault which is present in any sampler accompanies the probability that that same fault is a good indication of the merit of a sampler as a whole; and *vice versa* to what extent a low degree of agreement between judges as to the amount of a certain fault which is present indicates that that fault is not a good index of general merit in sewing. Table XV shows this correlation to be .89. This seems surprisingly high. It might at first thought seem to be explained by the fact that disagreement between judges which would make a low $q_1\ q_2$ coefficient of correlation would introduce an element of chance error into the measures of the amount of these faults about which judges were especially disagreed, and that this element of chance in the measures would, of course, lower the correlation between amount of that fault and general merit—the action of chance inaccuracies of a measure being always to lower any correlation based upon it. Therefore we should naturally expect to find a rather high positive correlation between the two sets of correlations themselves. This explanation would fully account for the average of .89 between the two columns of correlations, were it not for the fact that the correlation formula which we used does away with all attenuation of the coefficient of correlation which is due to chance inaccuracies of the paired measures. This being the case, another explanation of the coefficient of correlation of .89 must be sought.

Two possible explanations are worth considering. One is that people may agree more among themselves about the things that strike them as significant than they do upon other points. Table VIII, which shows the facts considered significant by eleven judges in evaluating samplers for general merit, does not seem to indicate that the points upon which these judges especially based their judgments, as far as they recognized what these were, are much the same as the points upon which judges agree, which is shown by high correlations in the $r\ q_1\ q_2$ column of Table XIV. This comparison is rather difficult to make, due to the difficulty of understanding just what is meant by many of the entries in Table VIII.

Another possible explanation is that the six judges who determined the value of the sixty-four samplers as to general merit prob-

ably disagreed among themselves in respect to the same faults that the sixteen judges who judged for each fault disagreed about. The six judges, each of whom presumably was influenced somewhat, at least, by all of the same factors which are definitely mentioned in our list of twenty-three faults, made *total* judgments depending upon these twenty-three features. The features upon which they agreed would, however, in the final estimate of them all, have more weight than those upon which they disagreed in determining the total amount of general merit. This is true because the influence of chance would bring about zero influence to the factors upon which they disagreed, since positive judgments and negative judgments would occur equally often concerning them. The final estimate of amount of general merit made by several persons would thus depend more upon the amount of those factors concerning which the judges agreed than upon the amount of those factors concerning which they disagreed. This probably is the proper explanation of the high coefficient of .89 for the correlation between reliability of judgment concerning a fault and its agreement with general merit.

It remained to find the correlation which existed between certain faults and certain others. Many such might be found of interest for various reasons. The author was especially interested in finding the correlation of the following five pairs of faults: Fault 1 and 2, "Stitches and spaces too large," with "Unevenness in size." This correlation, when corrected for attenuation, was found to be .88. This and the other correlations to be given here between pairs of faults were obtained from measures of only thirty-two samplers, instead of the sixty-four used for our former correlations; therefore, they have a somewhat smaller amount of reliability. The corrected coefficient of correlation between Faults 1 and 3, "Stitches and spaces too large" and "Line of sewing crooked," is .82; between Faults 3 and 23, "Line of sewing crooked" and "Individual stitches in backstitch and combination crooked," .91; between Faults 6 and 7, "Slant of stitches incorrect" and "Stitches not parallel in slant with one another," .99; between Faults 8 and 9, "Knots and fastenings large, conspicuous, and badly made" and "Knots and fastening insecure or absent," .39.

As stated earlier, we cannot be sure that judgments concerning the various faults were uninfluenced by other factors present in the sampler, except that all the judges apparently tried to eliminate all else and believed they had succeeded. We also stated that in

case they did not succeed in this elimination it is quite conceivable that there were constant errors working in both directions which would balance one another. If the judges succeeded in being uninfluenced by other factors, or if the constant errors balanced one another, the results give a true estimate of the amount of agreement which exists between these various pairs of faults.

The reader is again reminded that *all* of the correlation found to exist between any pair of faults is not by any means dependent upon these faults alone. Many other factors, such as general intelligence, general motor control, ability to pay attention, interest in sewing and in success itself, and the like, all enter in to swell the correlation, probably to an enormous amount. The method of Partial Coefficients of Correlation alone could eliminate such extraneous influences. Due to the many common causes, the presence of many of which are necessary for all efficient living, positive correlations, often of a high amount, have repeatedly been found by psychologists to exist between all desirable traits. Such positive correlations were found by McCall[21] to exist between all desirable traits,[22] though it is interesting to note that in many parts of the same subject, McCall did not find such high correlations to exist as we have found between various sewing abilities. Addition, he finds, correlates with problem solving, .32; visual vocabulary with reading, .82; copying addresses with handwriting, .52. Rogers, in the study of mathematical ability, quoted earlier, finds that algebraic computation correlates with matching equations and problems, .82; with reasoning, .44; with arithmetic problems, .50; with Trabue language scales, .41; and with Thorndike reading tests, .38. These latter two she finds correlate with one another, .87. Winch,[23] in a study of school children, found that the correlation between substance memory and imagination is .75. Hollingworth, in an earlier quoted study of disability in spelling, finds that the correlation between difficulty in recall and in recognition, both of the spelling of words, was .33. Seashore,[24] in his tests of musical ability, found the highest correlation of pitch discrimination, out of

[21] McCall, W. A., *Correlation of Some Psychological and Educational Measurements*, 1916.

[22] With the possible exception of traits involved in cancellation tests, if such traits are indeed desirable.

[23] Winch, W. H., "Some Relations between Substance Memory and Productive Imagination in School Children," *British Journal of Psychology*, vol. iv, pp. 95–125.

[24] Seashore, C. E., *Studies in Psychology*, University of Iowa, 1918.

fourteen tests, to be that with tonal imagery, .52. The highest correlation he found between any of these tests, was that between singing interval and singing keynotes, which gave a correlation of .61. To be sure there seems to exist an exception to the fact that desirable traits are positively correlated in the case of some features of the ability to draw. Ayer[25] studied the matter, and reports that he finds practically zero correlation between drawing and description and also between drawing and diagramming. The coefficients found were .023 for the former pair and —.052 for the latter. Diagramming and description he found to have a correlative coefficient of .231. Drawing, however, is an exception to this general truth, as has been noted by other authors besides Ayer.[26] It remains true that positive correlations between desirable traits are the rule, the extent of the correlation in each case depending upon the number of outside elements common to the two functions measured, and the degree of relatedness between the two functions themselves.

As we have stated before, knowledge of the correlations existing between various phases of one subject should be had for three reasons: (1) because they emphasize the fact that *in the main* such correlations are positive (this fact seems to be particularly true in sewing); (2) because they emphasize the fact that such correlations are seldom, if ever, 1.00, thus suggesting the necessity of individual analytic records; (3) because they show the *relative* difference between correlations of some elements of the total ability with that ability as a whole, and some other elements with the total ability, thus indicating which elements are in themselves most significant of the total ability.

The amounts of the actual correlations which we found between various pairs of faults are interesting mainly in a relative way. They show, for instance, that "larger size" of stitches slightly more often accompanies "unevenness in size" than it does a "crooked line" of sewing; that teachers who do not differentiate between "wrong slant" and "lack of parallelness in slant" do not thereby do as much harm as teachers who fail to call the child's attention to the difference between "crooked line of sewing" and "crooked individual stitches," since the latter two are less well correlated with each

[25] Ayer, F. C., *The Psychology of Drawing*, 1916.

[26] Albein, *Der Anteil der nachkonstruierenden Tätigkeit des Auges und der Apperception an dem Behalten und der Wiedergabe einfacher Formen*, 1907, p. 33. Terman, "Genius and Stupidity," *Pedagogical Seminary*, 1906. Rouma, *Le Langage Graphique de l'Enfant*, p. 199.

TABLE XVI. *Average amount of the fault 'stitches too large' for each of the six different stitches on sixty-four samplers, as judged by eight individuals, and the average for the six stitches. Also amount of stitches too large when judged for all stitches together by eighteen individuals*

Identification No. of Samplers	Basting	Running	Combination	Hemming	Backstitch	Overcasting	Average of Six Stitches Judged Separately	Six Stitches Judged Together for Stitches too Large
3	1.1	1.0	1.9	4.1	1.9	2.9	2.1	2.3
9	.9	1.6	2.1	3.1	4.0	3.6	2.5	2.3
11	1.2	3.1	3.2	3.2	2.6	.9	2.4	2.2
44	1.6	1.5	1.4	3.1	1.5	2.7	2.0	1.4
59	.9	4.4	3.5	4.5	3.5	.9	2.9	3.0
125	.6	1.2	2.4	1.0	1.8	2.9	1.6	1.4
137	1.4	3.0	3.4	3.9	2.7	1.2	2.6	2.5
204	.7	3.1	3.4	1.5	2.0	.6	1.9	2.3
242	.5	2.7	2.5		2.0	1.4	1.8	2.4
245	.7	1.7	.1	3.2	3.0	.9	1.6	2.6
258	1.2	3.2	1.7	3.7	1.9	1.2	2.1	2.9
308	1.4	1.9	1.4	4.6	2.	1.1	2.1	2.9
352	.7	.2	.4	1.2	.4	.6	.6	.5
367	1.5	1.7	2.0	2.1	1.1	1.4	1.6	1.6
374	1.6	.5	.7	.5	.9	.6	.8	.5
377	.9	.7	1.2	.2	.9	2.2	1.0	1.3
382	1.1	.5	.2	.5	.6	.2	.5	.3
415	1.2	1.9	2.2	2.5	2.6	.7	1.8	1.7
416	1.0	.2	.6	2.4	.5	.9	.9	.9
427	1.4	2.7	2.0	3.1	1.1	3.0	2.2	2.5
442	1.2	3.6	3.9	3.9	2.5	.6	2.6	2.9
452	.2	3.1	3.1	3.9	.9	1.1	2.0	2.6
494	.5	2.7	3.0	3.2	1.8	.6	2.0	1.7
499	.7	1.8	3.4	2.0	1.8	1.2	1.8	1.6
590	1.2	.2	1.0	.9	.4	1.7	.9	.9
634	1.2	2.4	3.2	3.1	3.1	2.1	2.5	2.9
665	.4	.1	2.5	.2	.1	1.2	.7	.5
692	1.2	.9	2.5	2.4	1.8	1.9	1.8	2.0
734	.5	1.0	1.6	1.9	1.0	1.9	1.3	1.3
746	.7	.6	1.0	1.5	1.0	1.6	1.1	.9

TABLE XVI—*Continued*

Identification No. of Samplers	Basting	Running	Combination	Hemming	Backstitch	Overcasting	Average of Six Stitches Judged Separately.	Six Stitches Judged Together for Stitches too Large
751	2.2	1.4	3.0	1.1	.7	3.4	2.0	2.1
841	2.2	1.0	1.5	1.9	.6	2.1	1.5	1.9
17	.1	.7	1.6	1.6	3.4	.7	1.3	1.3
33	1.7	2.7	2.1	3.0	2.1	3.6	2.5	3.1
48	1.9	1.7	2.1	1.5	.5	.5	1.4	1.8
92	.9	2.9	3.9	3.2	2.2	3.2	2.1	3.3
112	.6	.9	1.6	1.5	.4	1.5	1.1	1.2
114	.5	3.1	3.2	2.1	.5	2.6	2.0	2.1
138	1.6	2.4	1.5	3.9	1.2	2.2	2.1	2.0
146	1.2	2.1	2.6	2.0	2.0	1.6	2.0	1.9
212	.1	.4	1.4	.9	.6	.6	.7	1.1
244	.0	.9	1.0	.7	1.2	.5	.7	.8
247	1.4	3.1	4.2	2.4	2.7	1.4	2.5	2.9
275	.9			1.5	1.5	1.7	1.4	2.5
302	.5	.1	.2	.0	.5	1.0	.4	.4
329	1.1	2.5	.9	.4	.1	1.4	1.1	.8
336	.5	.9	.5	1.9	.4	.6	.8	1.1
377	1.1	.1	1.2	.2	.5	.2	.5	1.0
412	2.2	2.5	1.1	1.7	.9	.9	1.5	1.9
443	2.4	4.2	1.4	3.7	2.1	1.5	2.5	2.6
444	.5	3.7	3.4	3.2	2.4	1.5	2.4	2.3
447	.4	3.0	1.7	3.9	1.9	.9	2.0	2.1
581	.6	.4	.4	1.4	.4	1.4	.8	.8
605	.4	1.9	.5	2.6	.2	1.9	1.2	1.6
613	1.6		4.4	4.6	2.1	1.1	2.8	3.0
636	1.0	.6	.6	1.0	1.6	1.4	1.0	1.9
639	1.2	2.1	2.0	3.4	.9	3.5	2.2	2.9
644	1.1	1.6	1.1	1.5	1.1	3.0	1.6	2.6
646	2.0	1.2	1.0	3.1	.6	2.6	1.7	2.8
662	.87	.1	3.2	1.2	.0	2.1	2.5	1.4
668	.37	.9	2.0	.2	.4	1.4	1.4	1.1
694	.37	2.0	2.4	2.4	1.2	1.2	2.1	2.6
741	2.9	.4	2.7	2.0	.5	1.0	1.6	1.5
747	1.1	.1	3.0	.9	.1	.9	1.0	1.2

other than are the former pair. "Knots and fastenings unduly large or badly made and conspicuous," one might think, would correlate even negatively with "Knots and fastenings insecure and absent." It is quite conceivable, however, that a large and conspicuous knot may be an insecure one; however, the chances are that the attempt to make the less detailed list of faults mutually exclusive, was not successful, and that part of the positive correlation of .39, which exists between these faults, and which is not dependent upon the outside factors which we have so often mentioned, is due to the overlapping of the terms "badly made" and "insecure."

The study of the twenty-three faults which has just been completed could be carried still further. The detailed analysis of possible faults given in Table X indicates that most of the twenty-three faults here studied could themselves be subdivided in very much the same way as faults in general were analyzed and studied above. As a sample, we have analyzed some of the subdivisions of Fault 1. This fault is "Stitches and spaces too large." In the detailed analysis, reference to spaces is omitted. Six subdivisions of the fault were studied, namely, stitches too large in basting, running, combination, hemming, backstitch, and overcasting respectively.

In this new test eight women acted as judges. They went through the same set of sixty-four samplers which had been judged for the twenty-three faults. This time they passed judgments on the six faults just mentioned, going through all of the sixty-four samplers and working for one of these faults before considering another. They were in ignorance of their rating for any other fault when rating a sampler for an additional fault. The same precautions and directions for marking were given them as had been given to the judges of the twenty-three faults. The possible scores they might give the samplers ranged again from 0 to 5. It was again emphasized that the entire range of possible scores need not be covered for every fault. Zero in every case should mean that no amount of the fault was present. Table XVI gives the average amounts of the scores given by the eight individuals for each sampler in each of these six faults. The average of the six faults is given, and the score for the average amount for the different samplers of Fault 1, "Stitches and spaces too large," from our earlier set of twenty-three faults, is repeated here from Table XII for comparison. Table XVII gives the same data in a different form. This table is comparable to

TABLE XVII. *Distribution of amounts of stitches too large for different kinds of sewing. The 25 percentiles, medians and 75 percentiles are indicated by connecting lines. Data from Table XVI.*

Amount	Basting	Running	Combination	Hemming	Backstitch	Overcasting	Average of Six Stitches Judged Separately	Six Stitches Judged Together for Stitches too Large
.0	1			1	1			
1	2	5	1		3			
.2	1	3	2	4	1	2		
.3								1
.4	5	3	2	1	6		1	1
.5	(7)	2	2	2	(6)	2	2	3
.6	3	(2)	2		4	7	1	
.7	5	2	1	1	1	2	3	
.8							3	3
.9	6	5	1	3	5	(7)	2	3
1.0	(2)	3	4	2	2	2	3	1
1.1	6		(2)	1		3	(3)	3
1.2	(9)	2	2	(2)	3	5	1	(2)
1.3							2	3
1.4	4	1	4	1			3	3
1.5	1	1	2	6	2	3	2	1
1.6	4	(2)	3	1	1	2	5	3
1.7	1	3	2	1		2	(1)	2
1.8		1			4		4	1
1.9	1	3	(1)	3	3	3	1	(4)
2.0	1	1	4	(3)	(4)		7	2
2.1		2	3	2	3	(3)	(6)	3
2.2	3		1		1	2	2	1
2.3								4
2.4	1	2	2	4	1		2	1
2.5		2	3	1	1		6	(3)
2.6			1	1	2	2	2	5
2.7		(4)	1		2	1	1	
2.8								1
2.9	1	1				2	1	6
3.0		2	(3)	1	1	2		2
3.1		5	1	5	1			1
3.2		1	4	(5)		1		
3.3								1
3.4			4	1	1	1		
3.5			1		1	1		
3.6		1				2		
3.7		1		2				
3.8								
3.9			2	5				
4.0					1			
4.1				1				
4.2		1	1					
4.3								
4.4		1	1					
4.5				1				
4.6				2				
4.7								
4.8								
4.9								
No	64	62	63	63	64	64	64	64
M	1.0	1.6	1.9	2.0	1.1	1.4	1.7	1.9
Q	.35	.55	.95	1.	.75	.6	.5	.65

Table XIII, which gave the same information concerning the twenty-three faults which is here given for the six faults of "Stitches too large" in six different kinds of sewing. Inspection of this table is equally valuable. It reveals at a glance the median and variability of this fault as it exists in the six different stitches, showing us here, through the presence of a high median, which of these faults is hard to overcome. We thus see that hemming and combination stitch have the worst records in regard to large stitches, whereas basting and backstitch have almost the same small median of 1. and 1.1 respectively. That the more general fault of "Stitches and spaces too large" has a median amount value equal to that of combination stitch (1.9) and just .1 less than that for hemming, which has the highest median amount of all these six faults, indicates that in scoring for the more general fault judges were more influenced by the stitches which had comparatively large amounts of the fault "Stitches too large" than by those which had little of it.

As was done in the case of the twenty-three faults, so now for the six kinds of "Stitches too large," the correlation was found between each one and general amount of the whole sampler. This was found as before for the sixty-four samplers. The four coefficients of correlation necessary to correct for attenuation and the final corrected coefficient of each of the six faults with general merit are given in Table XVIII.

A comparison of Tables XVII and XVIII shows that hemming, the stitch in which this fault of large stitches is hardest to overcome, as shown by the high median amount of it (2.0) which existed in our sixty-four samplers, is also the stitch which, in respect of this fault, has the greatest amount of significance for general merit. Overcasting and basting stitches being "too large," are faults rather easy to overcome, and with a small amount of significance for general merit. This we should expect from the fact that these two stitches often are purposely made rather large. However, running-stitch is not, and yet we find that in regard to fault in size of its stitches it is approximately half way between these two and hemming.

On the other hand, backstitching is shown to have a different amount of importance by the two tables. Table XVII shows by its low median value of 1.1 that it is comparatively seldom present in large amounts. In spite of this, however, it is seen by Table XVIII that the importance of large stitches in backstitching, as a symptom of general merit, is fairly great, the two having an inverse correlation

TABLE XVIII. *Coefficients of correlation found between general merit of the whole sampler, and the fault 'stitches too large' as it occurs in each of six kinds of stitches*

p_1 = general merit as judged by three judges using the scale
p_2 = general merit as judged by three other judges using the scale
q_1 = amount of the fault 'stitches too large' for each kind of stitch, as judged by four judges
q_2 = amount of the fault 'stitches too large' for each kind of stitch, as judged by four other judges

Kind of stitch being judged for the fault 'stitches too large'	$r p_1 p_2$	$r q_1 q_2$	$r p_1 q_2$	$r p_2 q_1$	$r pq$
Hemming	.967	.809	−.734	−.699	−.81
Running	.967	.965	−.675	−.787	−.76
Backstitch	.967	.869	−.678	−.642	−.719
Combination	.967	.881	−.511	−.487	−.541
Overcasting	.967	.787	−.175	−.268	−.249
Basting	.967	.779	−.211	−.078	−.148

of .72. Reversely, combination stitch has a less important significance for general merit (an inverse correlation between them is .54), but it is much more often present in large amounts, the median amount of this fault possessed by sixty-four samplers being 1.9.

Column $r\, q_1\, q_2$ of Table XVIII gives the correlations between random halves of the eight judges concerning "Stitches too large," in the case of the six stitches. These eight judges agree well as regards this fault when it appears in running stitch (the reliability coefficient here is .96). When it appears in overcasting and basting they agree only to the amounts represented by reliability coefficients of .79 and .80. When judging for the same fault in hemming they agree very little better. The reliability coefficient for this stitch is .81.

Similar studies of the detailed analysis of the twenty-two remaining faults should be made which would give information of a kind comparable to that given here for parts of Fault 1 (not all the parts, for "Spaces too large" was not measured in the detailed study). Only by means of such studies is a teacher enabled accurately to know in what stitch or part of the product various kinds of faults are apt to be found, and only so is she enabled to know how much of a symptom of general merit a certain fault may be, according to the particular stitch in which it is found.

SECTION V

THE RELIABILITY OF MEASUREMENTS OF SEWING

RELIABILITY IN RELATION TO THE KIND OF STITCH

In order to determine the importance of each of the five sewing stitches represented upon the sampler in indicating general merit of the sampler as a whole, 100 samplers which had already been evaluated by twelve judges were used. These samples were cut into five pieces each in such a way that the different kinds of sewing were separated. The basting stitches were ripped out, and the backstitching and overcasting respectively were ripped from the two pieces into which the original seam was cut. It was unfortunately impossible to use all of each of those stitches without destroying the seams and making the judgments upon them somewhat ambiguous. The scheme for giving identification numbers to the different samplers, was such that it would be impossible from them alone to identify the parts which originally formed one whole, unless one were familiar with the numbering key. We thus had 100 samples each of the following stitches: hemming, running, backstitching, overcasting, and combination stitch.

The 100 samplers were made in one school by fifty children, each child having made two samplers. This fact is not of importance here, but will be dealt with later. All that is necessary to note here is that the selection was made in a way to produce a fairly normal surface of distribution.

Twenty judgments were made upon each of the five sets of 100 samples of sewing. Directions were given that the 100 pieces be placed in seven piles in accordance with their merit as samples of hemming, overcasting, running, etc. The differences in merit between these samples in the seven piles were to be equal.

'For each of the five stitches the judgments of ten alternate judges were grouped and served as one measure of merit for that stitch. The judgments of the other ten judges served as the second measure. The correlation between general merit of the original sampler before cutting and the merit of each of the five stitches separately found after cutting was obtained; also the correlation between the sum

of the scores for the stitches judged separately and total merit before cutting. The use of two measures for each fact made it possible to free the correlations from the attenuation which would be found in a raw correlation obtained from only one measure of each fact, which attenuation is due to chance inaccuracies of the

TABLE XIX. *Coefficients of correlation between samples of separate stitches and the sum of these stitches with general merit of the whole sampler, giving all the coefficients necessary to correct for attenuation*

p_1 = one measure of total merit which is the average of judgments made by six individuals

p_2 = one other measure of total merit which is the average of judgments made by six other individuals

q_1 = one measure of the samples of each stitch, which is the average of judgments made by ten individuals

q_2 = one other measure of the samples of each stitch, which is the average of judgments made by ten other individuals

$r p_1 p_2 = .88$

Stitches correlated with general merit	rq_1q_2	rp_1q_2	rp_2q_1	rpq
Hemming	.88	.78	.69	.82
Running	.82	.67	.70	.80
Combination	.91	.62	.66	.71
Backstitch	.95	.61	.62	.66
Overcasting	.92	.62	.50	.61
Sum of six stitches	.96	.87	.85	.94

original measures. The intercorrelations between the two measures are also useful in furnishing the means of testing the reliability of the judgments. Table XIX gives the actual coefficients of correlation used in obtaining the final coefficients, corrected for attenuation. Column $r\ q_1\ q_2$ contains the coefficients of reliability of the judgments made upon these samples of sewing. Column $r\ p\ q$ contains the corrected coefficients of correlation between general merit and the several stitches, as judged by the samples of stitches which were used. These coefficients constitute a measure of the indication which the sample used of each sewing stitch is of general merit of the sampler as a whole. Since the amount of sewing upon which the judgments were based was only half as much in the case of the backstitching and overcasting, half of the original sewing in each of these cases having been ripped away, as has

already been explained, there is a possible insecurity in drawing conclusions from our data so far as concerns comparisons of these two stitches and the other three. The correction for attenuation, however, probably puts all the inter-comparisons of the corrected coefficients on the same footing since they give the correlations

[1] We cannot be positive that the one and one-half inches of overcasting and backstitch were indeed absolutely random samples of the whole amount. Therefore, in order to discover whether a difference in the amount of sewing represented by the overcasting and backstitching samples (one and one-half inches each) and that represented by the samples of the other three stitches which were twice as large, was really sufficient to cause a difference in the judgments made upon them, ten new judgments were made upon one-half of each of the 100 samples of hemming, running, and combination stitch, the other half being covered by a piece of paper or cloth. The correlation was then found, for each of the three stitches, between the average of these ten new judgments which were based upon one and one-half inches of sewing and the average of ten of the twenty judges who had three inches of sewing to base judgments upon. These correlations in the case of each stitch were compared with those which had been found between the two groups of ten judges, each of which group had judged upon the basis of three inches of sewing. Table XX shows both these correlations, the difference between the correlations, and the P.E. of this difference. In the case of two of the stitches (hemming and combination stitch) the first of these correlations (that between judgments all based upon three inches of sewing) is higher, and the difference between the correlations is more than twice the P.E. of the difference. This indicates that a difference possibly does exist between judgments based upon one and one-half inches of this kind of sewing and judgments based upon three inches, but it can hardly be said that it is proved that this difference exists, since due to chance alone 17.7 cases out of 100 would fall beyond twice the P.E. of the difference, were there in reality no difference at all in the facts judged. In the case of the running stitch the second correlation is higher and the difference between the two being less than twice the P.E. of the difference, it is probable that one and one-half inches of running stitch yield the same judgments that three inches do.

TABLE XX. *Coefficients of correlation between average of ten judgments made upon three inches of sewing of various stitches and average of ten other such judgments; and between the average of the first ten, and the average of the ten judgments based upon one and one-half inches of sewing of the same stitches.*

q_1 = average of first ten judges upon three inches of sewing
q_2 = average of second ten judges upon three inches of sewing
q_3 = average of ten judges upon one and one-half inches of sewing

	Hemming	Running	Combination
rq_1q_2	.882 ± .015	.824 ± .022	.911 ± .011
rq_1q_3	.768 ± .007	.854 ± .018	.832 ± .021
Difference between above correlations	.114	—.03	.079
P. E. of difference	.034	.056	.043

which would be found if we had had an infinite number of judges, instead of ten or twenty, rate the specimens. With an infinite number of judges the result from rating one and a half inches taken by chance from a specimen would presumably be the same as that from rating the whole of it.[1]

The last column of Table XIX shows that the five stitches vary, as indicators of general merit, from .61, which is the correlation of general merit with overcasting, to .82, which is its correlation with hemming. In deciding upon the general merit of an article of sewing, other things being equal, one should give more weight to the hemming stitches and the running stitches (the correlation of running with general merit being .80, almost as high as that of hemming). Least weight of all these five stitches should be given to overcasting, and almost as little to backstitch, combination stitch taking an intermediate place. These different weights should be assigned to the different stitches, not because they each contribute such-and-such proportional amounts to general merit as a whole; in fact, what the independent contributions are of each stitch to general merit is not shown in Table XIX. That matter is suggested just below where the partial coefficients of correlation of each stitch with general merit are given. Even, however, with only such knowledge at hand as is given in Table XIX, that is, knowledge of the actual correlations which exist between each stitch and general merit, we may say that more weight should be given to the hemming than to other stitches, because it better indicates general merit, this being shown by its higher coefficient of correlation with it.

The partial coefficients were then found of each stitch with general merit, when the effect of the other stitches was eliminated.

TABLE XXI. *Correlations found between each stitch and every other stitch, as well as between each stitch and general merit*

	General Merit	Combination	Running	Overcasting	Hemming	Backstitch
General merit		.71	.80	.62	.83	.67
Combination	.71		.78	.37	.58	.62
Running	.80	.78		.42	.67	.64
Overcasting	.62	.37	.42		.35	.26
Hemming	.83	.58	.67	.35		.47
Backstitch	.67	.62	.64	.26	.47	

For a discussion of this method and its implications the reader is referred back to Section 4. Table XXI gives the intercorrelations of each stitch with every other stitch, a necessary preliminary knowledge to the finding of the partial coefficients. Table XXII gives the partial coefficients of each stitch with general merit.

TABLE XXII. *Partial coefficients of correlation between general merit and each stitch. Each coefficient was found in two ways and the average taken*

	Combination	Running	Overcasting	Hemming	Backstitch
First finding	.0885	.2902	.6666	.7496	.4912
Second finding	.0843	.2902	.6729	.7519	.4897
Average	.0864	.2902	.6698	.7507	.4904

RELIABILITY IN RELATION TO SYNTHETIC VS. ANALYTIC JUDGMENTS

In order to compare synthetic judgments, or those made upon the samplers as a whole, when these were judged for general merit before being cut, with the analytic judgments made upon the cut-up portions, the correlation (corrected for attenuation) was found which exists between the total value of the sampler before cutting, and the sums of the value of the five different parts when these were judged separately. This correlation is given in Table XIX. It may seem strange that it is not 1.00 rather than .94. Several reasons besides chance may account for this. First, all the basting stitches were removed from the cut-up pieces, and half of the overcasting and backstitching. A probably more important omission from the 'parts' of that which was contained in the 'whole', was the factor of relative arrangement of the stitches upon the total sampler.

Another factor which, if it existed, would surely lower the correlation between any whole and the sum of its parts, even if all of these were surely included, would be that in adding together the parts, a wrong weighting might be attached to each, as far as its relative importance in determining the whole was concerned. The weighting assigned here was that of giving equal emphasis to every stitch. That the stitches have not exactly equal importance as symptoms of general merit has been shown. Everything considered, it is really rather surprising that the correlation is so high as

.94, and that it is so seems to indicate that the five different stitches must have very nearly an equal effect in indicating general merit.

It is of importance to the theory of educational measurement to know how reliable are judgments made upon total merit of some school subject, such as a bit of penmanship, an English composition, or a simple sewing sampler, when compared with the sum of judgments made upon various features of the same product. The use of data obtained from Table XIX enables us to answer this question in regard to the sewing sampler. We find there in the line for "Sums of five stitches," that $r\,p_1\,p_2$ is .88 and $r\,q_1\,q_2$ is .96; $r\,p_1\,p_2$ is the correlation between six judges and six other judges for total merit of the whole sampler; $r\,q_1\,q_2$ is the correlation between the sum of ten judgments upon each of the five stitches separately and the sum of ten other judgments upon the same five stitches separately. By using the formula[2]

$$r_x = \frac{n\ r_1}{1+(n-1)\,r_1}$$

we can find what the comparison between these correlations of $p_1\,p_2$ and $q_1\,q_2$ would be if they had been made upon the basis of the same number of judgments. Thus, by substituting 10/6 for n, we find that the correlation which we would obtain for $r\,p_1\,p_2$ would be .92 if we had used twenty judges instead of twelve. The comparison of reliability properly then should be made between coefficients of .92 and .96 respectively for general merit judged *in toto* and for general merit when judged by the sum of the elements which compose it. When we consider the increase of time and energy necessarily spent in evaluating separately for five stitches, as compared to a rough general rating in terms of general merit alone, it is evidently a much wiser educational practice to use total merit as a criterion rather than the sums of judgments made upon the various stitches separately, since the coefficient of reliability is almost as high for general merit judged as a total.

RELIABILITY IN RELATION TO THE NUMBER OF SAMPLES

To determine the reliability of one or more samples of a child's sewing in evaluating her real ability in that kind of sewing, the data described under A of this section were used, but instead of treating the 100 total samplers before cutting and the 100 samplers

[2] Suggested by William Brown in *The Essentials of Mental Measurement.*

of each of the five separate samples of sewing obtained after cutting as forming each one set of facts to be measured, the 100 pieces of sewing to be measured in each of the six cases were, for this portion of our experiment, separated into those which were made first and those which were made one week later by the fifty children who produced them. Correlations were then obtained between the

TABLE XXIII. *Coefficients of correlation corrected for attenuation, between the first and second performance, in various kinds of sewing, of fifty pupils; also the number of samples, such as were used, of the various kinds of sewing which would be necessary to produce reliability coefficients of .90, .95, and .975.*

	Correlation between first and second performances	Number of samples necessary to produce a correlation of .90 with as many more samples	Number of samples necessary to produce a correlation of .95 with as many more samples	Number of samples necessary to produce a correlation of .975 with as many more samples
Total sampler	.72	4.3	9.2	18.9
Backstitch	.80	2.3	4.9	10.0
Running	.64	5.0	10.6	21.8
Hemming	.62	5.5	11.6	23.8
Combination	.60	6.1	12.9	26.4
Overcasting	.40	13.6	28.7	59.0

first and second performances of the total sampler, and of each of the five stitches separately. Table XXIII gives these coefficients of correlation, corrected for attenuation.

It is important to note the very great difference which exists between some of these correlations. A child's performance at overcasting, made one week later, has a correlation with the earlier performance of only .40; whereas the correlation between two samples of backstitching made at the same time as the overcasting and upon the same samplers, is .80. The other three stitches are more alike in this respect, the correlations between the two performances being respectively .64, .62, and .60 for running, hemming, and combination stitch.

These correlations furnish a certain indication of the reliability of one sample of each of these stitches as a measure of a child's real ability in that line. We proposed to go beyond this and to determine how many samples of each stitch would be necessary in order to know that we had certain specified degrees of probability that a child's real ability to make this stitch was measured. This was done by the use of the following formula:[3]

$$r_x = \frac{n\ r_1}{1 + (n-1)\ r_1}$$

As the formula stands, the problem is to find the correlation which would obtain if some different number of measures of the original facts to be correlated were used. For instance, if we knew the correlation which exists between one measure of a pupil's ability to solve addition problems and another measure of this same ability, and if we desired to know the correlation which would be found if the average of two such measures were correlated with the average of two others, we would find it by substituting 2 for n in the formula. We, however, made use of the formula in a different way. Our problem was—knowing r_x as representing a certain specified degree of probability that the ability in question was truly measured, to determine the number of measures of each fact which, when correlated with an equal number of such measures, would result in the desired measure of probability (that is, in r_x). This required number was then found by solving the above formula for n. In other words, having already found, as shown in the first column of Table XXIII, the correlations which exist between one performance and one other such performance of the whole sampler and of each stitch separately, we now desired to find how many of such performances we would need to measure in order to obtain correlations of .90, when these were averaged and correlated with the average of an equal number of such measures. The second column of Table XXIII gives the required number of samples necessary to produce correlations of .90 when paired with as many more. The difference noted above between overcasting and backstitch is here brought out more emphatically. The results show that 13.6 samples of the latter stitch yield no more reliable information concerning a child's real ability in producing it than do 2.3 samples of backstitch.

[3] Suggested by William Brown in *The Essentials of Mental Measurement.*

In each case that many samples of each stitch, if correlated with as many more such samples, would yield a coefficient of correlation of .90. In the third and fourth columns of the same table are given the number of samples of the total samplers and of each stitch which would be necessary in order that correlations of .95 and .97½, respectively, would obtain, if an equal number of such samples were correlated with them.

The two performances of each child's sewing, which we used for the basis of the previous discussion, were made in two successive sewing periods with an interval of one week. We desired to find the improvement made by the children as a whole of the second performance over the first, in the case of the samplers when judged for general merit before being cut, and in case of each of the five stitches. The improvement in the total sampler and in each stitch was found by taking the difference between the average according to the first ten judges (six only for the total sampler) for the samples of all the first performances and the average for the samples of all the second performances according to the same ten (or six) judges. This gave us one measure of difference. In order to test its reliability the σ of the difference was found by the formula in Whipple's *Manual of Mental and Physical Tests*:[4]

$$\sigma \text{ improvement} = \frac{\sqrt{\sigma 1^2 - 2r\sigma 1\,\sigma 2 + \sigma 2^2}}{\sqrt{n}}$$

We thus had one measure of improvement and one measure of the unreliability of this improvement for each kind of stitch. A second measure of each was found in exactly the same way, except that the scores of the other ten (or six) judges made upon the same samples of sewing were used. The average between the two measures of improvement and the average between the two measures of σ were taken as the final measures for these two facts for each of the five stitches. In all cases, except hemming, the σ of the improvement was about one-third the improvement. For hemming it was a little more than half the improvement.

Table XXIV shows the amounts and σ's of improvement. As it happens, this improvement in case of the total sampler and each of the stitches is negative. The amount, however, is very small.

[4] Page 27.

TABLE XXIV. *The improvement of the second sewing performance over the first, which was made in the previous sewing period*

	Total Sampler	Over-casting	Combi-nation	Back-stitch	Hem-ming	Run-ning
Improvement according to first ten judges (only six for total sampler)	−3.00	−5.6	−5.5	−6.1	−2.9	−6.5
σ of improvement according ing to first ten judges (only six for total sampler)	.89	1.8	1.4	1.2	2.0	1.6
Improvement according to to second ten judges (only six for total sampler)	−4.4	−4.4	−6.2	−4.5	−2.3	−6.8
σ of improvement according to second ten judges (only six for total sampler)	.81	2.0	1.5	1.2	1.2	1.5
Average between two previous measures of improvement	−3.7	−5.0	−5.8	−5.3	−2.6	−6.6
Average between two previous measures of σ	.85	1.9	1.5	1.2	1.6	1.6

Possibly the second sampler may have been made more rapidly or upon a bad day, or there may have been a dislike to repeating the task. No sewing practice in school intervened between the two performances. The instruction given to the supervisor in charge of the class was that the children make the two samplers under conditions as nearly as possible identical.

What is of more interest than the fact of a very small amount of negative improvement, is the *relative* difference in this negative improvement as shown in the five different stitches. Hemming is especially different in this respect from the other four stitches. The fifty children who made each two samplers were in this stitch much more constant, a fact which is brought out clearly by both sets of judges. The average deterioration in their samples of hemming was 2.56. In the running stitch the deterioration was greatest, being 6.66. (The unit in which these measures are expressed equals one-tenth of a step between successive piles of the seven into which the 100 samples of each stitch were sorted by the twenty judges.)

RELIABILITY IN RELATION TO THE CHARACTER OF THE JUDGES

In the making of the scale various classes of persons acted as judges. Following are facts which were gathered at that time concerning relative reliability between these different judges.

When thirty-six judgments in all had been made upon 129 or 854 samplers there was a chance to compare the amount of agreement existing between the average of various groups of judges and the average judgment of the thirty-six. The groups compared were one of eight past or present sewing teachers; one of eight women who said they knew considerable about sewing, but who had never taught it; one of eight women who said they knew very little about sewing; and one group of eight men, all of whom claimed that they knew nothing about sewing. The remaining four persons in the group of thirty-six who were not included in the four groups of eight, were all women who had a fair amount of knowledge of sewing.

Table XXV shows the correlations between each of these groups and the group of thirty-six, and also the correlation of half of each of these groups with the other half. The most interesting feature of this table is the indication that sewing teachers agree less with the group as a whole than do any of the other three groups, including the men. If the sewing teachers, while differing from the group of thirty-six, had yet shown a greater amount of agreement between themselves than do the other groups, we might think that they were the most expert judges, but since they agree among themselves no better than do the other two groups of women, we hardly feel justified in saying that they are any better judges of sewing. One constant difference between their judgments as a group and that of the other groups, is that they place the samplers on the average a little lower, or nearer the zero end of the scale. That this is true is seen by comparing in Table XXV the average position assigned samplers by the various groups, the average being here expressed in terms of the original ten to twelve piles into which the samplers were sorted.

One possible source of error in concluding from Table XXV that sewing teachers are less reliable judges of sewing than are other women competent in other respects, is the fact that six of the eight sewing teachers made judgments upon 854 of the samplers instead

TABLE XXV. *Coefficients of correlation, showing agreement of four groups with the average of thirty-six judges, and the agreement of one half of each group with the other half; also average position assigned samplers by each group*

	Sewing teachers	Women who know about sewing	Women who do not know about sewing	Men
Coefficient of correlation between each group and average of thirty-six judges	.911 ± .01	.986 ± .002	.991 ± .003	.978 ± .003
Coefficient of correlation between one half of each group and the other half	.958 ± .005	.957 ± .005	.966 ± .004	.78 ± .023
Average position assigned samplers	6.18	5.99	5.82	5.87

of 129. This was a tedious task and, we might well think, one which would prevent the exercise of one's best judgment. However, four of the group of women who knew much about sewing were also among the original twelve judges, three of them having judged 1,184 samplers and one 854; whereas only one sewing teacher judged the full number of 1,154 samplers. Of the group of women without knowledge of sewing two made judgments upon 1,184 samplers; the others upon only 129. The fact also that the correlation between six and six judges upon 854 samplers is as high as .93 makes us feel that the comparatively low correlation between eight sewing teachers and the group of thirty-six individuals which they partly compose is not due to the fact that a larger number of them than any other group judged a large number of samplers.

An odd feature of Table XXV is that the correlation between the group of eight men and the total group of thirty-six is as high as .978; whereas the correlation between one half of the men and the other half is as low as .78. The author thinks that the explanation is probably due to the fact that two of the men who happened to be grouped in the same half when the intercorrelation was found, made extremely poor judgments, little if any better than chance would allow. These men were students in a class in Experimental Psychology. Probably the task of arranging sewing samplers was one which did not call forth their best efforts. The other six men evidently made conscientious judgments which, on the whole, compare favorably with those made by any of the women.

RELIABILITY OF JUDGMENT IN RELATION TO THE DISTRIBUTION AND TO THE KIND OF PRODUCTS JUDGED

That a small distribution of the qualities of the product to be judged makes for less reliability in the judgments passed upon them is shown by comparing two correlations which were found in the course of our previous investigation. In making the scale it was necessary to find the reliability of the first twelve judgments made upon the 854 samplers. The correlation between six of these judges and six others was found to be .933 with a P.E. of .003. Later it was necessary to find the correlation between judgments of these same two groups of six judges each upon 100 only of the 854 samplers. This correlation is given in column $r\ p_1\ p_2$ of Table XIX. It is .88. In this latter case of 100, both very good and very poor samplers were omitted, and this probably accounts for the correlation of .88 as compared to the one of .93 when a larger range of samplers was judged.

That reliability of judgments depends upon the kind of products judged has been shown in previous tables.

In column $r\ q_1\ q_2$ of Table XIX are given the correlations between six individuals and six other individuals upon 100 samples of five different stitches. These correlations show that the reliability of judgments passed upon certain stitches is much greater than is the reliability of judgments upon certain other stitches. To be more specific, these correlations show that judges agree most concerning backstitch (reliability coefficient, .95) and least concerning running (reliability coefficient, .82). For overcasting, combination stitch, and hemming, there seems to be about an equal amount of reliability.

It has also been shown in Section 4 that reliability of judgments which are passed upon various faults is dependent greatly upon the particular fault judged. In the $r\ q_1\ q_2$ column of Table XIV are given the coefficients of correlation between eight judges and eight others upon sixty-four samplers judged for each of the twenty-three faults studied. An inspection of that column gives useful information as to which faults people agree most about. "Stitches not parallel in slant with one another" seems to be most easy to identify, as judged by the high correlation .94 between random halves of judges, whereas "Stitches and spaces too large," which one might think was equally identifiable, has a correlation of only .77 between random halves of the judges.

RELIABILITY IN RELATION TO THE USE OF THE SCALE

In column $r\ p_1\ p_2$ of Table XVIII is given a correlation of .97 between three and three judges for general merit of the sixty-four samplers which were judged also for faults. The correlation of .88 of Table XIX, which was between judgments passed upon the 100 samplers made by fifty children from one school, is a correlation between six and six judges. This larger number of judges in itself should make for greater reliability. The fact that the correlation here is only .88 as compared with the correlation mentioned above of .97 is partly due to the fact that it is concerned with samples which cover a smaller range of general merit, but probably it is due still more to the fact that the judgments in this case were made without the aid of the sewing scale described in Section 3, and in the case of the correlation for the sixty-four samplers the sewing scale was used. One purpose of our scale is that by its use variability of judgments will be lessened. The comparison between these correlations seems to show that it fulfills its mission.

In order to further test the usefulness of the scale in reducing variability between judges, a new experiment was undertaken. Seven sewing teachers from Washington Irving High School, New York City, and one sewing supervisor of Teachers College were the subjects. One former sewing teacher from another city was included in the group by making two of the judgments recorded. These sewing teachers were asked to grade certain articles in two ways. Half of them first went through all on the familiar basis of 100 per cent. The other half first made judgments by the aid of, and in terms of, the scores of our scale. The two halves of the judges then went through all the articles again, grading them by the method they had not already used. Not all of the sewing teachers graded all the articles in both ways, but for each article there were at the very least, four judgments made according to each method of judging.

The articles judged were five of our samplers, five aprons, five bags, and five samples each of small pieces containing only hemming and overcasting, respectively. The author was somewhat doubtful about the feasibility of using the scale for judging one stitch only, but the results seem to indicate that such is possible. Since our aim is to compare the variability between judgments made by the different individuals in the two ways of judging, the average devia-

TABLE XXVI. *Comparison between the variabilities of sewing teachers from the same group in judging the same products, when using the per cent. method of grading, and when using the sewing scale*

Product judged	Judged by Per Cent Method				Judged by Sewing Scale			
	Median per cent. assigned by from four to six sewing teachers	Average deviation from median	σ of distribution of medians for each kind of product judged	Average deviation ÷6σ of group	Median scale value assigned by from four to six sewing teachers	Average deviation from median	σ of distribution of medians for each kind of product judged	Average deviation ÷6σ of group
Hemming			25.2				28.0	
No. 7	30	18		.11	79	7.5		.04
No. 13	20	14		.09	55.5	9.0		.05
No. 21	60	14		.09	99.5	14.0		.08
No. 42	85	8		.05	131.0	3		.01
No. 90	75	10		.06	131.0	6.75		.04
Average				.08				.04
Overcasting			17.4				19.9	
No. 20	35	24		.22	96.5	18.12		.15
No. 23	60	23		.22	131.2	3.5		.02
No. 32	50	15		.14	117.5	6.5		.05
No. 34	85	17		.16	150.5	8.5		.07
No. 92	85	4		.03	147	9.25		.07
Average				.15				.07
Samplers			23.1				35.5	
No. 245	30	14		.10	67	4.4		.02
No. 367	70	10		.07	131	9.4		.04
No. 452	20	9.6		.06	51	8.6		.04
No. 746	80	8		.05	143	4.4		.02
No. 841	60	15.4		.11	104	10.4		.04
Average				.07				.03
Aprons			21.5				31.8	
No. 11	20	12.5		.09	47	11.25		.05
No. 12	55	20		.15	72.5	28.25		.14
No. 13	80	7.5		.05	137	6		.03
No. 14	26	9.66		.07	57.5	19.75		.10
No. 15	40	15.4		.11	64	6.0		.03
Average				.09				.07
Bags			18.0				36.0	
No. 16	45	13.75		.12	47	4.75		.02
No. 17	20	10		.09	43	5.66		.02
No. 18	35	7.5		.06	43	6.40		.01
No. 19	86	7		.06	137	3		.02
No. 20	47½	9		.08	67	6.33		.018
Average				.08				.017
Average of averages				.094				.045

tion of the judgments from their median was found in each case for each article. Table XXVI gives these average deviations from the medians for each article judged, both by the use of the scale and by the per cent. method of grading. The scale values, before the judging was made, were multiplied by ten to eliminate the use of the decimal point. Obviously, since the scale values thus run up to a higher point than do the per cent. values (100 being, of course, the highest possible value in the latter case, as 164 is the highest possible value by the scale), thus making a greater range in the former case in which variability may occur, some account should be taken of this fact in comparing the amounts of average deviation in the two cases, for when the median is higher and the range is greater, other things being equal, the variability also will be greater. In order to eliminate this error the following procedure was adopted. Each group of five kinds of one article was treated separately, and the σ of the distribution of the five articles (each article being scored by the median of the judgments passed upon it) rated in all five of these groups was found for each method of judging. Six times the σ of any distribution, containing as it does 997 cases out of 1,000, represents for all practical purposes the range of that distribution. In order, then, to equate the average deviation from the respective medians of the judgments made by the two methods, these average deviations were in the case of each article divided through by the range for that kind of article, as represented by six times the σ of the distribution. This was done for each of the two methods of scoring. Six times the sigma was used instead of the exact range because it is a measure less subject to inaccuracies caused by extreme cases.

An inspection of Table XXVI shows that in the case of almost every separate article, and at any rate, in the case of the average of every kind of article, even the gross variability between judges is greater when the per cent. method of judging was used. When the gross amounts of variability are divided through as just indicated by 6 σ of the distribution of each kind of article in the case of each of the methods of judging, the difference between the variability when scale and per cent. method are used respectively is even more striking. This striking difference in the variabilities again goes to show that an objective standard to which all persons may refer does eliminate much of the personal equation of judges, which personal equation leads them in the case of the per cent. method of

marking to disagree very widely among themselves. It must be remembered, too, that most of the judges in this case were teachers from the same school and, at that, from a school with a particularly high reputation for its sewing instruction. The per cent. standard among these teachers, who meet together frequently for conference, is surely very much more uniform than it is for sewing teachers in general.

It must also be recognized that these teachers were used to the per cent. method of grading and vastly preferred it to the scale method, which method they were using for the first time and which, being unfamiliar, seemed very burdensome to them. This fact would certainly not make them especially careful in their gradings by the scale. Both their unfamiliarity with it and their dislike of a new method would be factors which would probably act to lessen the reliability of their judgments made by the scale at that time. In spite of these facts, we see from Table XXVI the great advantage of judging by the scale. When the average deviation of the judgments upon each article is divided through by six times the σ of the distribution of that kind of article and the average of all these average deviations is found, it is .94 when the per cent. method is used and .456 when the scale is used. If the test had been made by judges familiar with the use of the scale, and more kindly disposed to using it, and by judges from different schools whose standards in terms of per cent. would thus be less uniform than they were among the group we measured, the difference in their variability when judging by the two methods would probably be even more pronounced.

It will now be worth while to study Table XXVI with the idea of evaluating the different amounts of relative variability in the judgments made by the two methods of grading when passed upon the five different kinds of articles.

As has already been said, the author was somewhat doubtful about the feasibility of using the scale for judging one stitch only, and included samples containing only overcasting or hemming stitch as an experiment. Table XXVI shows, however, that the comparative difference between variability of judgments made by the two methods of scoring is about as great for hemming and overcasting as it is for all the articles taken together.

That the judgments made upon the aprons when the scale was used were, on the whole, more variable than judgments upon any

other article, except the samples of overcasting, was probably due to the fact that aprons 11, 12, and 14 were unfinished. The judges were asked to disregard this fact and grade them upon the basis of the completed portions, but this direction was apparently not followed by all. Apron 15 is made almost entirely by machine, and contains very little hand work. It was included simply as an experiment, the author not believing that the scale gradings of such an article would yield such reliable judgments as were found.

Probably the bags furnish the best test of what the scale is really intended to measure. They are completed articles made by hand, not by machine, and they contain most of the stitches illustrated in our samplers. Table XXVI shows that the variability between judges in grading bags by the use of the scale is .018. This is less than their variability in judging any of the other articles. It is very considerably less than the variability of judgments upon bags made without the use of the scale. In fact, this difference between the variabilities according to the two methods when bags alone are judged, is much greater than the difference between the variability according to the two methods of judging when we consider all five kinds of articles judged.

Our conclusions earlier were based upon the difference between the variability for the two ways of judging, based upon the average variabilities for the five articles used in our experiment. We have shown that many of these articles were introduced merely to test the *possible* use of the scale for sundry purposes. The *main* use of the scale, however, is to measure sewing merit in articles similar to the bags, representing as they do completed useful products in which most of our scale stitches and no other stitches are represented. This being the case, we may say that when the scale is used for such a purpose its value in reducing variability of judgments is even greater than we have so far claimed.

SECTION VI

A SUMMARY OF RESULTS

The following tasks have been accomplished, or problems solved by this investigation:

1. A scale has been made for measuring merit in certain elements of hand sewing. This scale consists of half-tone reproductions of fifteen sewing samplers so reproduced and arranged that three illustrations of each sampler represent all of the sewing upon the originals. The sewing consists of a hem, a seam, sewed with a backstitch (or stitching stitch), whose raw edges are overcast; and a few inches of running and of combination stitches. The range of general merit in the sewing upon these samplers is from that of an imbecile boy with very poor motor control and eye defect, to that of the most expert sewer the author could find among several colleges of practical arts. Between these two extremes are the reproductions of thirteen intermediary samplers. All fifteen of the samplers have defined values at approximately equal steps. That this sewing scale is of value in greatly reducing the variability of judgments made upon articles of sewing has already been proved. Its further usefulness in other ways seems to be fully promised by the knowledge we have of the part which other scales have recently played in educational practice.

2. An estimate has been found of the number of samplers, similar to those described above, which it would be necessary to have made by a child in order that her real sewing ability may be measured with various degrees of accuracy. If it is desired that our knowledge of her sewing ability have a correlation of .9 with her real sewing ability (that ability which is the average of *all* her sewing) it is necessary that from four to five such sewing samplers be made by her on different occasions, and the average of all of them be used as the measure of her ability. If we desire accuracy to the extent of 95 per cent. correlation with a perfect measure, between nine and ten such samplers would be required to produce it. For accuracy represented by 97½ per cent. of correlation with that same "real" ability, practically nineteen such samplers would have to be made

by the child. Of course, if one were measuring a *group* of children, to find the average sewing ability of a class or school system, so many samplers would not be necessary. One sampler made by each child in a group of fifty children would furnish probably a more accurate measure of the class average than would four samplers made by one child furnish, as to her ability.

3. Analyses have been made of the faults which may be present in these same elements of hand sewing, which the samplers, above discussed, represent. The first analysis, a very detailed one, containing mention of about 100 faults, in a mutually exclusive list, was made primarily for the purpose of assisting teachers of sewing in clarifying their own and their pupils' thoughts, as to the exact lines along which improvement is needed. Pedagogically, it is recognized that this treatment would be bad if given to the child in this negative form. "Faults" rather than "excellences" have formed the basis of this part of our research, both because in practice one is more often struck by the presence of the former than of the latter, when evaluating sewing productions, and also because many pairs of two faults are associated with only *one* excellence, which is a golden mean between the two. The detailed analysis is believed to be useful because it may help the teacher to think analytically when such thinking would be helpful to her in her teaching. It is not believed that analytical thinking such as it suggests would be helpful at all times.

The second analysis, which consists of twenty-three faults in a mutually exclusive list, was made partly to supplement the former one, it being wiser upon some occasions to use a somewhat less detailed scheme in thinking of the faults of sewing. It was made also as the basis for measuring the faults which actually do exist in children's sewing.

4. The relative amounts of faults which are actually present in children's sewing when these are expressed in terms of our less detailed list of twenty-three, have been found. The two faults which appear in the greatest amounts are first, "Unevenness in size," and second, "Individual stitches crooked in combination and backstitch." Large amounts of the other faults appear in the following order: "Stitches not parallel in slant with one another," "Slant of stitches incorrect" appears in the same amount as "Stitches and spaces too large." "Line of sewing crooked or not parallel with edge of material" appears in the same amount as the two following:

"Edges not trimmed," and "Knots and fastening large, badly made, and conspicuous." Next is "Bad arrangement of stitches on samplers." "Hem badly turned in" appears in the same amount as does "Threads left hanging" and "Broken and soiled and double thread used"; next is "Wrinkled material," which appears in the same amount as does the average of all the twenty-three faults; then "Seam or hem too large"; "Knots or fastenings insecure or absent" is equal in amount to "Basting at wrong distance from edge of hem or seam." "Stitches unsuited for purpose for which intended" is next, followed by equal amounts of "Soiled material," and "Edges of seam unevenly joined." In decreasingly small amounts appear "Stitches drawn too tight," "Evidence of ripping and threads of material pulled," "Stitches not drawn tight enough," "Seam or hem too small," and "Extra material caught in under sewing."

These relative amounts of the actual existence of the faults have been compared with two other facts which we have determined concerning them. One is the amount of significance which each fault has as an indicator of general merit. The other is the relative amount of agreement which exists between individuals in estimating the amount of these various faults present in a given case. These two latter facts, the significance of a fault as an indicator of the amount of general merit present, and the relative agreement of judges concerning faults, were found to be similar; that is to say, it was found that faults which were conspicuous as being good indicators of general merit were, on the whole, faults concerning which there was a high degree of agreement between judges, and *vice versa*, that those faults which were poor as indicators of general merit, were faults concerning which judges were apt to disagree. The correspondence, however, was not found to be so great between either of these and the amount of a fault actually present. We were particularly interested in a comparison between the amount of a fault's importance as indicating general merit and the amount of its actual existence. The order of the twenty-three faults as indicators of general merit is as follows: "Unevenness in size"; "Stitches or spaces too large"; "Incorrect slant"; "Unsuited for purpose intended"; "Line crooked and not parallel with edge of material"; "Stitches not parallel in slant with one another"; "Individual stitches in backstitch and combination are crooked"; "Hem badly turned in"; "Bad arrangement"; "Wrinkled material"; "Threads left hanging and broken and soiled"; "Edges of seam unevenly joined";

"Stitches too tight"; "Knots and fastenings large"; "Stitches not tight enough"; "Evidence of ripping and threads pulled"; "Basting at wrong distance from edge of hem"; "Extra material caught in sewing"; "Soiled material"; "Seam or hem too small"; "Seam or hem too large"; "Knots and fastenings insecure or absent"; "Edges not trimmed". If this order is compared with that given above, which refers to the *amount* of faults present, the reader will see that they are somewhat similar, but that it is very unsafe to infer from one order what the other would be. The safe practice would be to learn both orders. In this way, one would always know that "Unevenness in size of stitches and spaces" is a fault present in large amounts, (therefore one probably overcome only with difficulty) and that it is also a fault, the amount of whose presence is a strong indication of how much general merit in sewing it accompanies. On the other hand, one would know that "Stitches too tight" is a fault which happens seldom, large amounts of it being comparatively rare, yet one which is rather important as an indicator of general merit; and so on through the list.

In general, we may say that our results have shown that various traits in sewing, as indicated by our list of twenty-three faults, are highly correlated—more so on the whole than are various subdivisions of most elementary school subjects. This indicates that, on the whole, ability to sew is general rather than specialized, that whatever outside influences make for success in one phase seem also to operate to a rather large extent in most of the other phases. In spite of this general tendency toward high correlation there are, of course, cases of exception in individual pupils, the diagnosis of whose peculiarities will be greatly aided by the use of our two analyses. There are also various degrees of the amount of correlation which exists between the different phases of this ability. The most important ones to note are those which have already been suggested in the list of indicators of general merit. The order in which the faults appear in the list is, of course, the order of the amount of correlation which each one has with general merit. These correlations vary from .94 to .19, and it is evidently unsafe to infer high correlation between *all* elements of sewing ability, even though, as stated above, we find such correlations to be particularly high in sewing. The safe practice is to know the general rule of high correlation, the individual correlations found between each trait and

general merit, and the idiosyncrasies of the special class or individual pupil with whom we are to deal.

5. Five fundamental stitches of hand sewing—hemming, overcasting, backstitch, running, and combination stitches—have been studied in the following ways (the results of the various findings concerning the different stitches are brought together in Table XXVII):

TABLE XXVII. *Relative importance of five sewing stitches according to various ways of dealing with them*

	Agreement between ten judges and ten other judges concerning one hundred samples of each stitch	Correlation, corrected for attenuation, between one hundred samples of each stitch judged by twenty judges and general merit judged by twelve judges	Partial coefficient of correlation with general merit	Correlation corrected for attenuation between first and second performances of fifty children, as judged by twenty judges	Number of samples necessary to produce a correlation of ninety with as many more samples	Number of samples necessary to produce a correlation of ninety-five with as many more samples	Number of samples necessary to produce a correlation of ninety-seven and one-half with as many more samples	Improvement between first and second performance of fifty children, judged by twenty judges	Median amount of fault 'stitches too large' which occurred for each stitch in sixty-four samples	Correlation corrected for attenuation between amount of fault 'stitches too large,' as judged by eight judges, and total merit as judged by six judges using the scale. Sixty-four samples used	Agreement between four judges and four others, concerning the fault 'stitches too large' as it occurred in each stitch. Sixty-four samples used
Overcasting	.92	.61	.67	.40	13.6	28.7	59.	—5.	1.4	—.25	.79
Running	.82	.80	.29	.64	5.	10.6	21.8	—6.66	1.6	—.76	.96
Combination	.91	.71	.09	.60	6.1	12.9	26.4	—5.84	1.9	—.54	.88
Backstitch	.95	.66	.49	.80	2.3	4.9	10	—5.29	1.1	—.72	.87
Hemming	.88	.82	.75	.62	5.5	11.6	23.8	—2.56	2.0	—.81	.81

(*a*) The relative reliability of judgments has been found according to the stitch concerning which it is made. Backstitch is thus seen from the table to be the stitch concerning which judges are most agreed as to whether a certain sample of it is good or bad; running stitch is the stitch concerning which they agree least.

(*b*) The relative importance of each of these stitches as indicators of general merit has been found. Hemming is seen to be of most importance in this connection and overcasting of least

(*c*) The partial coefficient of correlation has been found of each stitch with general merit, the four other stitches having been

eliminated, the effect upon the correlation of their common elements thus having been removed.

(*d*) The reliability of one sample of a child's sewing in each of these stitches has been found, and the number of samples of each which would be necessary in order to know that there were various degrees of probability that her "real" ability in this kind of sewing was measured. These are the same kind of estimations which are reported above under 2 to have been made for the total sampler. The most interesting fact concerning the results now under discussion is the great amount of relative difference indicated between some of the stitches in this respect. To arrive at a measure of each stitch which would be equally accurate to that which would be rendered as to general merit by from four to five sewing samplers, it would be necessary to have only 2.3 samples of backstitch; hemming, running, and combination stitch would need to be represented by about as many samples as would be necessary for the total sampler, while to arrive at a measure equally reliable concerning a child's ability in overcasting, 13.6 samples of it would be required.

(*e*) In two successive sewing periods we found that a class of children, as a whole, did worse in the second period. Of course, such negative improvement occurring once has little significance for educational theory. We already know that improvement is never at an invariably uniform rate, but that unevenness in the practice curve due to chance is the rule, and that slight regression is of frequent occurrence. What is of much interest, however, in this connection is again a relative matter in connection with the different stitches. Table XXVII shows that three of the stitches were affected somewhat equally in this respect, the deterioration being about the same in combination, backstitch, and overcasting. In running stitch it is slightly greater. Hemming, on the other hand, shows a deterioration less than one-half that of the average of the other stitches.

(*f*) The fault "Stitches too large" which, as we have reported, had been measured along with twenty-two other faults, was measured in detail as it applied to the five different stitches under consideration here. Just as each of the twenty-three faults earlier had been measured in three respects, so this one in its detailed form as applying to the separate stitches was measured in three respects. These three measures, which are found in the three last columns of Table XXVII, give scores for (1) the amount of the fault which is present

in each stitch, (2) the amount of indication which the fault has for general merit, according to the stitch in which it appears, and (3) the reliability of judgments concerning the fault as it appears in the different stitches. The fault "Stitches too large" is greatest when it appears in hemming and almost as great when in combination stitch. Backstitch contains least of this fault. From the next to last column we find that hemming is also the stitch in which the fault is most serious; in the sense, that is, that "Stitches too large" in hemming is more of an indication of lack of general merit *in toto* than is "Stitches too large" in any other stitch. It must be emphasized, however, that this does not at all imply that *intrinsically* lack of the fault when in this stitch rather than in any other enters in more as a component of general merit. Whether it does or does not, it is impossible for us to know from the results we have at hand. Only by treating our results according to the method of partial coefficients of correlation, which we have not done in this connection, could we determine anything at all concerning the amount of contribution which any partial trait makes to a more generalized ability. When it was said above that *in a certain sense* one fault was more "serious" when present in hemming than in any other stitch, it was meant that it was so only because, as things were, the fault in hemming does have a higher correlation with general merit than the fault in any other stitch. What extraneous influences may enter to make this correlation so high, our results give no basis for estimation. The results which we found when partial coefficients of correlation were found between each stitch and general merit indicate that if the method had been applied to faults also, we would have seen that often very little of the actually obtained coefficients of correlations between general merit and any fault is really due to this one factor alone. As things stand, however, hemming is the stitch in which this fault, when it appears, has the highest correlation with general merit. The fault when in overcasting, on the other hand, has only a low correlation with general merit.

As to the reliability of judgments passed upon the fault, there is not a great deal of difference, according as it appears in different stitches, with the exception of the running stitch. Apparently judges are more agreed among themselves as to when "Stitches are too large" in this stitch, than when they are too large in any other.

APPENDIX I

A TEST OF TRANSFER FROM SEWING TO "NEATNESS" AND TO "ÆSTHETIC APPRECIATION"

Woolman, in a pamphlet entitled *A Talk with Mothers on the Value of Sewing Teaching*, says,"The German women are more apt to have clean, orderly homes than our American women, on account of the thorough drilling they have had in sewing."

To test properly the truthfulness of this and similar claims for the transfer value of the teaching of sewing, it would be necessary to conduct a carefully planned experiment of the usual type for testing transfer, using a practice group and a control group, testing both as to neatness in housekeeping both before and after the sewing practice of one group, and comparing the results of improvement of the two groups. Such an experiment would require much time and would be rather difficult to handle. Some light, however, may be cast upon the question of transfer by finding the correlation which exists between the two abilities of sewing and housekeeping. If children who learn to sew are thereby neater in their household duties, it would necessarily always follow that a high correlation would exist between these two abilities. If, therefore, we are able to obtain measures of these two abilities for some group of individuals and find a low correlation between them, the truth of the statements concerning large amounts of positive transfer is disproved. If, on the other hand, we find a high degree of correlation between the two abilities, the question of transfer is left open, unchallenged, but also still unproved, for positive correlation between two abilities may be due to many other causes than that of transfer from one ability to the other.

In order to obtain some evidence upon this subject, the author obtained from the Home Economics Division of the Iowa State College measurements upon "neatness" in certain fields of work for three different classes, one of twelve, the other two of fifteen students each. The measurements consisted of a rank order of arrangement of the students according to the amount of their "neatness" in the various subjects. The rankings were made by the

instructors in each course, the directions to these instructors being that each rank the students independently, and in ignorance of other rankings. One class was thus ranked for "neatness" in Millinery, Applied Dress Design, and Personal Appearance. The two other classes were each ranked for neatness in the following: Foods, House, Advanced Textiles and Clothing, and Personal Appearance.

TABLE XXVIII. *Coefficients of correlation, obtained by the rank method, found between 'neatness' in various fields for three college classes*

Subjects Correlated for Neatness	Class A	Class B	Class C
Millinery and applied dress design	.36 ± .161		
Millinery and personal appearance	−.53 ± .132		
Applied dress design and personal appearance	.03 ± .183		
Foods and house		.53 ± .147	.67 ± .101
Foods and advanced textiles and clothing		.49 ± .156	.17 ± .178
Foods and personal appearance		.32 ± .183	.73 ± .086
House and advanced textile and clothing		.74 ± .093	.27 ± .169
House and personal appearance		.71 ± .102	.71 ± .091
Advanced textile and clothing and personal appearance		.80 ± .072	.72 ± .089

Correlations were found by the rank-order method between each pair of measures for each class, and are given in Table XXVIII. Inspection of this table shows that some positive correlation, in the case of two of the classes, exists between neatness in all of the four fields. That it is of a degree to warrant the statements quoted at the commencement of this chapter, seems very doubtful, even if the negative records of Class A are disregarded. Some common factor may exist between neatness in these various fields. Just what this element is and how it should best be trained are important educational problems. That the most efficient way to foster it is to teach sewing seems surely not to be the case, for the correlations between subjects other than those three involving sewing ability (Applied

Dress Design, Textiles and Clothing, and Millinery) are quite as high or higher than are the correlations between these three subjects and others.

Another claim frequently made is that æsthetic appreciation is developed by learning to sew.

Thus, after describing the training of hand and eye which comes from learning to sew, Margaret Swanson in *Needlecraft in the School* (1916) says: "At twelve years of age the girl has acquired the power of hand and eye, needed for the relatively smaller, more finical seams of dolls' clothing. This possession of power implies a development of more or less artistic appreciation."

The facts to be reported, which throw some light upon this claim, were gathered by the author in connection with another research. Only that part of the experiment which helps to solve the present problem under consideration will be reported here.

This problem is that of measuring the transfer effect which sewing practice produces upon æsthetic appreciation.

The method of correlation which was used in connection with the study of the transfer value of sewing to neatness was used here also, the argument being in this case that if there is positive transfer from sewing to æsthetic appreciation it should show itself by the existence of a positive correlation between tests of sewing ability and tests of æsthetic appreciation.

The subjects of the experiment were twenty-two students from the senior class of the woman's department of the Carnegie Institute of Technology. Two measures of their sewing ability were obtained by using their ratings in "sewing" for the Plebe and Sophomore years.

A test in æsthetic appreciation in color combination in dress was devised in the following manner: Fifteen copies of the same dress design were colored by hand in fifteen different ways, different colors and combinations of colors being used. These were mounted upon a gray background. Thirteen women who were considered experts in the subject were asked to arrange these fifteen colored dresses in the order of their importance, beauty and suitability of the color or colors to the style of gown were both to be considered. Nine of these experts were teachers of dress design. That the test is a reliable one is shown by the fact that the correlation between the median position assigned to each dress by half of these judges and the median position of each dress according to the other half of the thirteen judges was found to be .93.

The same directions concerning the arrangement of the fifteen pictures of colored dresses were given to each of the twenty-two students in this class to be tested. The position in which each student put each dress was compared with the median position in which that dress was put by the group of thirteen experts. Two scores were found for each student. The first score was the sum [1] of the deviations between the positions into which she put each of the first eight dresses from the positions into which these same dresses were put by the median expert judgment. The second score was the sum of her deviations in respect to the other seven of the total number of fifteen dresses.

Having two measures of each of the abilities to be correlated made it possible to free this correlation from attenuation due to chance inaccuracies in the particular measures of the abilities which were used.

Letting p_1, p_2 and p represent, respectively, the two scores actually found for sewing ability and the real measure of that ability which an infinite number of scores would show, letting q_1, q_2 and q equal the same three measures of ability in æsthetic appreciation, the following correlations were found:

$$r\ p_1\ p_2 = .60$$
$$r\ q_1\ q_2 = .80$$
$$r\ p_1\ q_2 = -.24$$
$$r\ p_2\ q_1 = -.15$$
$$r\ p\ q = -.27$$
$$\sigma\ t.r. - obt.\ r = .136$$
$$\sigma\ t.r. - obt.\ r = .076$$

It is unfortunate that the number of students was so few. As a consequence, the unreliability of our correlations as shown for two of them by $\sigma\ t.r. - obt.\ r$ (true correlation − obtained correlation) is very great. No very great significance, therefore, can be attached to our findings, being subject, as they are, to much influence from extreme measures. The fact, however, that the $r\ p\ q$ or the correlation found here to exist between sewing ability and æsthetic choice is —.27, is probably a very strong indication that there is no great amount of *positive* transfer from sewing to the latter ability. As a matter of fact, the correlation being a minus one is probably due to the very erratic æsthetic choice of one student, who has

[1] Regardless of signs.

a rather high amount of sewing ability. Had this student been eliminated from the reckoning, a correlation of about zero would probably have been found.

Neither our test for transfer from sewing ability to "neatness" nor that for transfer from sewing ability to "æsthetic choice of color combination" is detailed or reliable enough to serve as a certain proof that there is no such transfer. They do, however, furnish evidence sufficient to make us say that there is a strong probability against any great amount of transfer from sewing to either of these two other abilities.

APPENDIX II

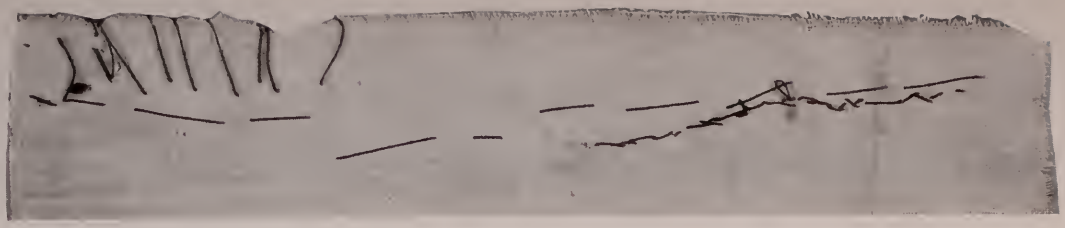
A SEWING SCALE

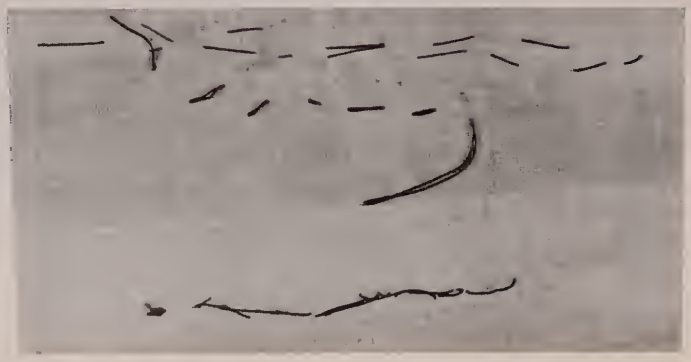
The Sewing Scale described in the foregoing pages is shown on the following inserts.

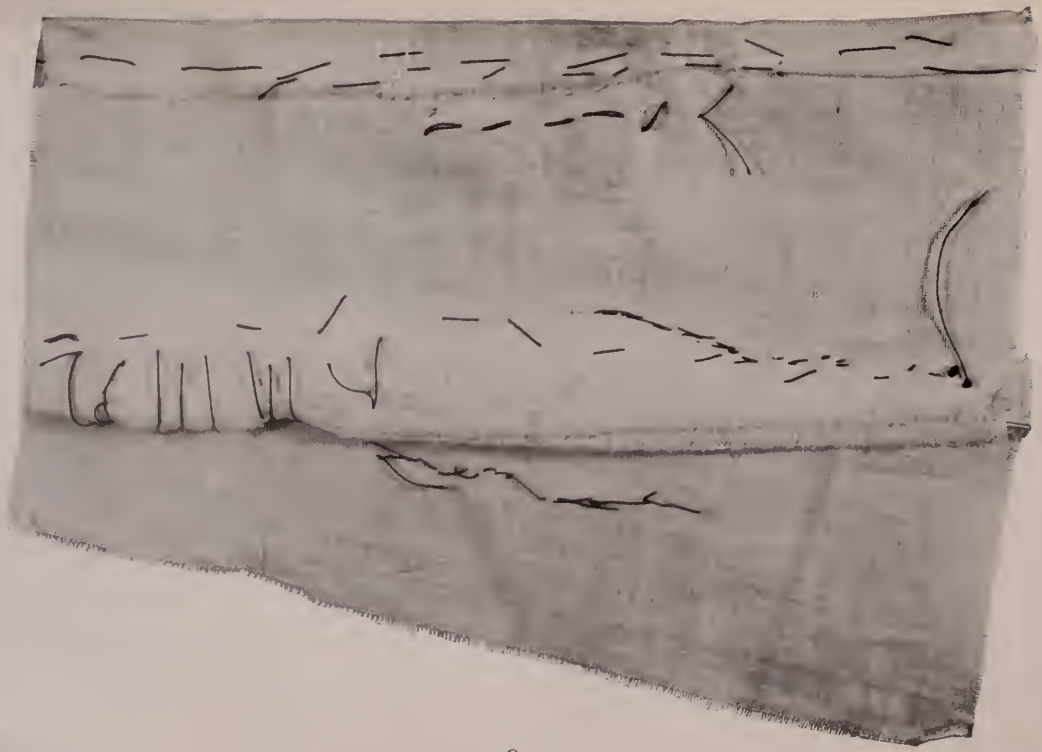
Each page shows three views of the same sewing sampler. One view gives the full-sized view, showing a hem, a seam, hemming, basting, running, backstitch, overcasting, and combination stitch. A second view shows the reverse side of the seam, and of the overcasting, basting, and backstitch. A third view shows the reverse side of the hem, of the hemming stitch, basting, running, and combination stitch. The numerical values of each entire sampler is printed beneath it.

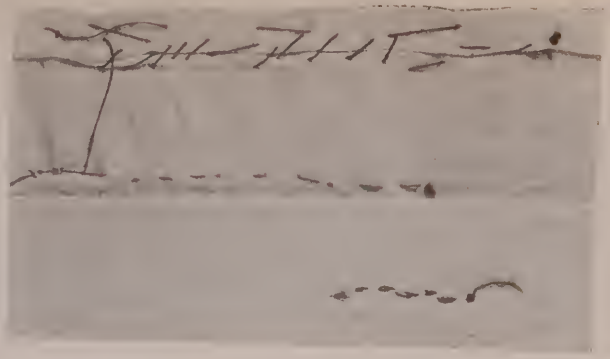
DIRECTIONS FOR MEASURING

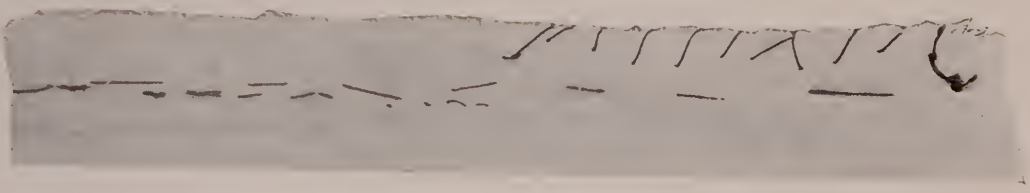
Compare the quality of the sewing to be judged with the quality of the samplers in the scale. Assign to the sewing you are judging the numerical value of that evaluated sampler which equals it in general merit. In case it falls between two scale samplers in merit, assign to it a numerical value between the values of these samplers, proportionate to its difference from each.

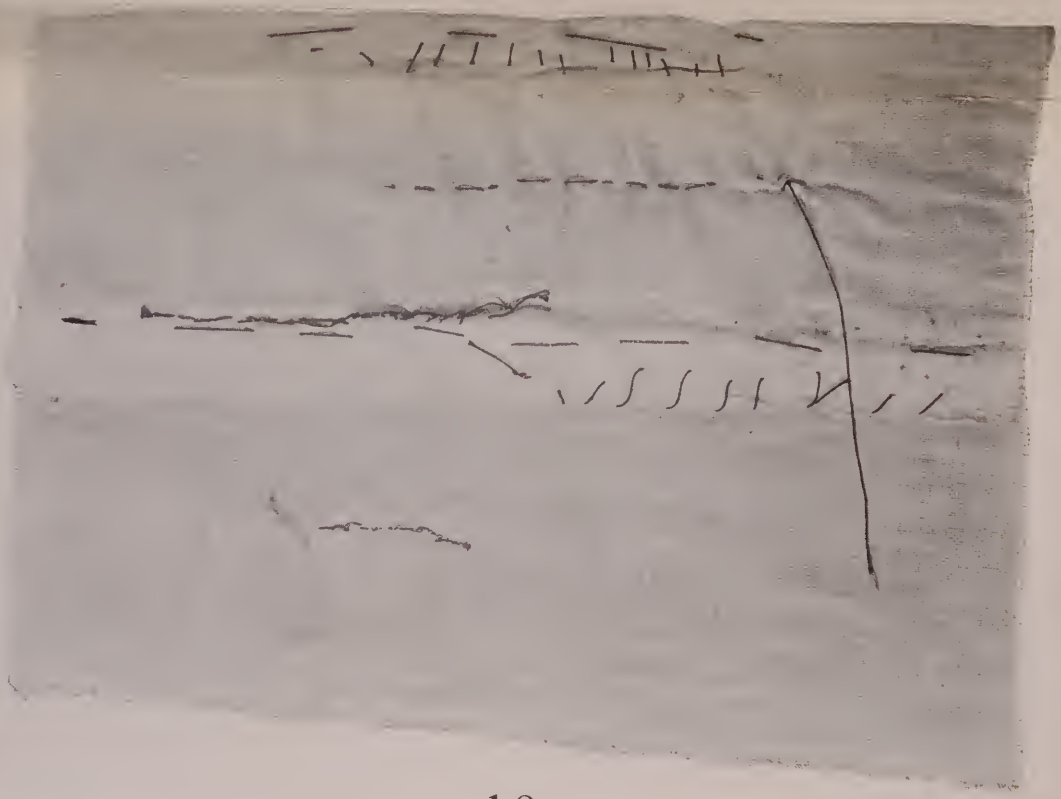


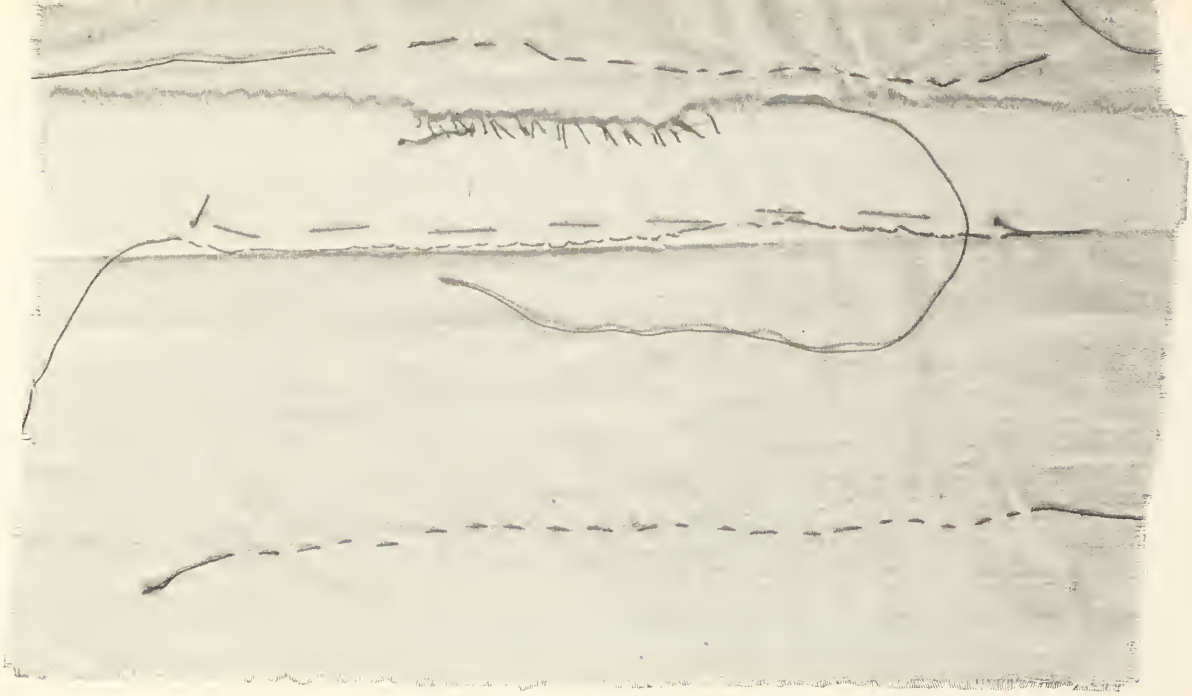




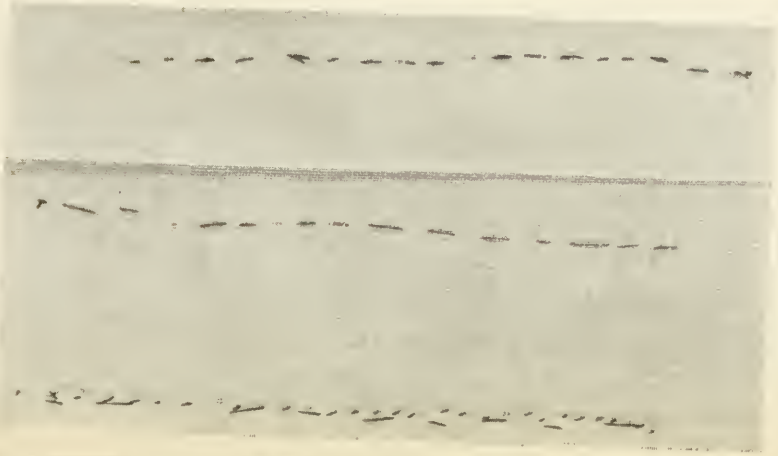


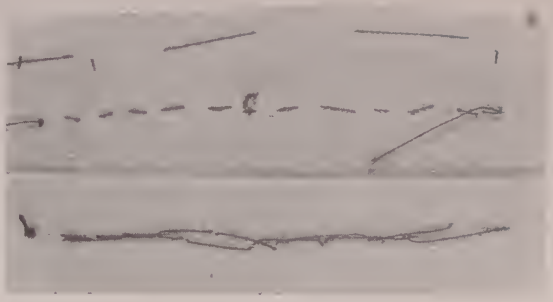


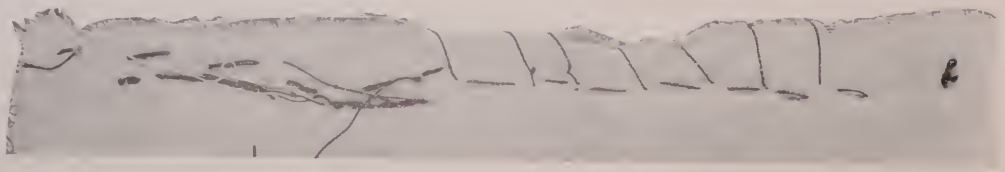


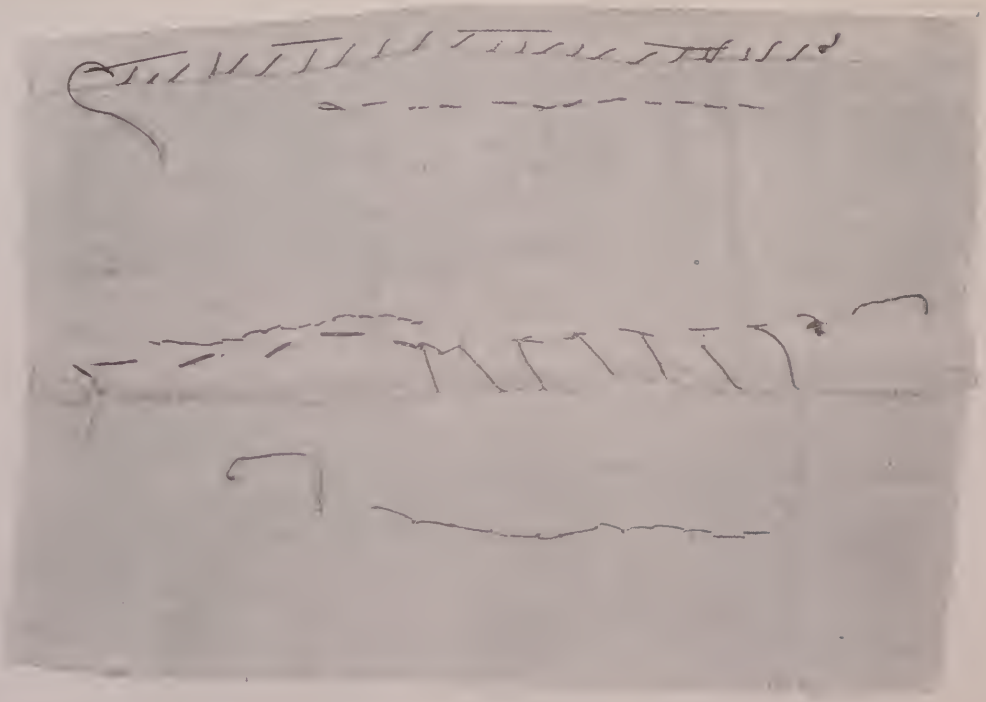


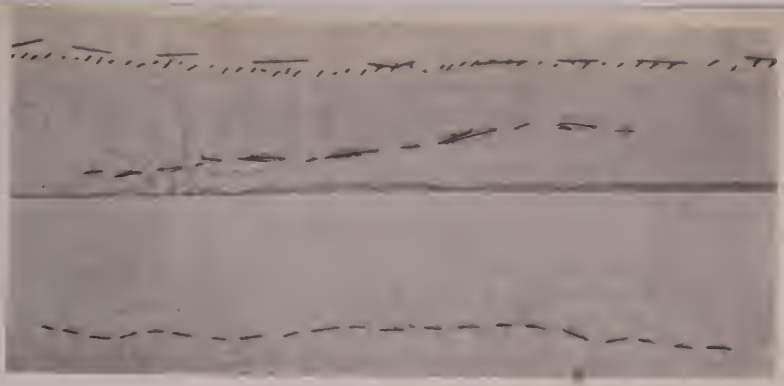
4.3

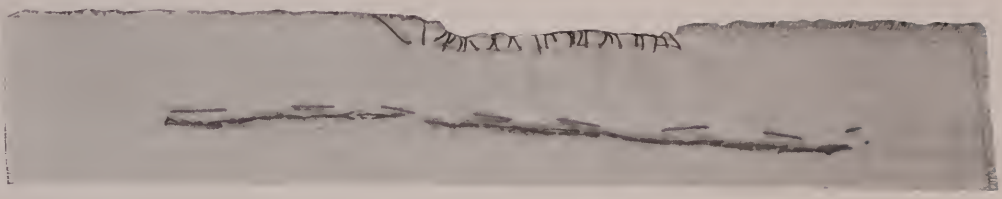


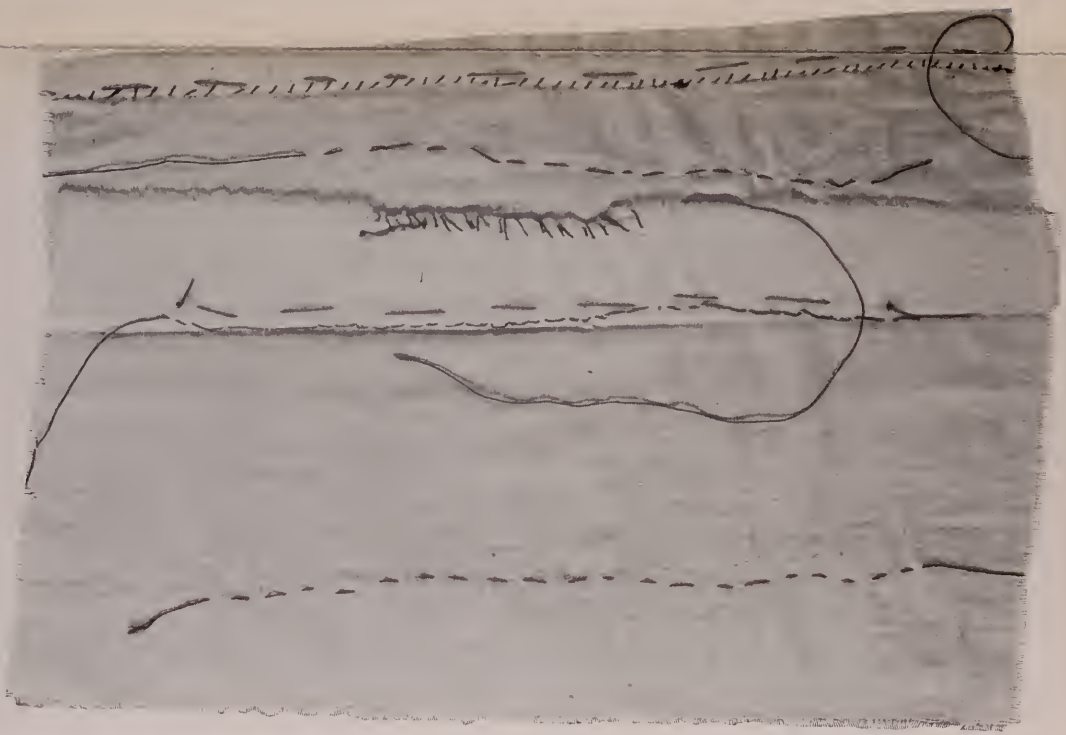




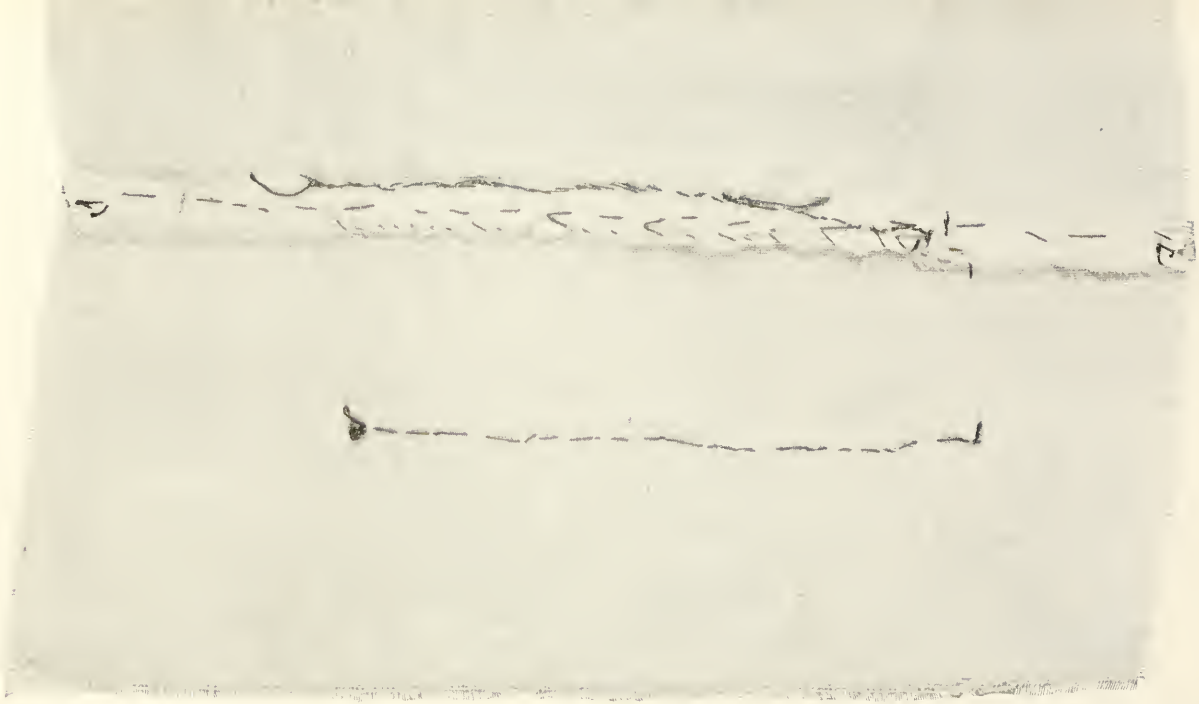




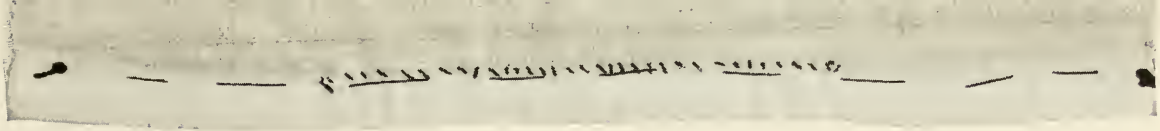


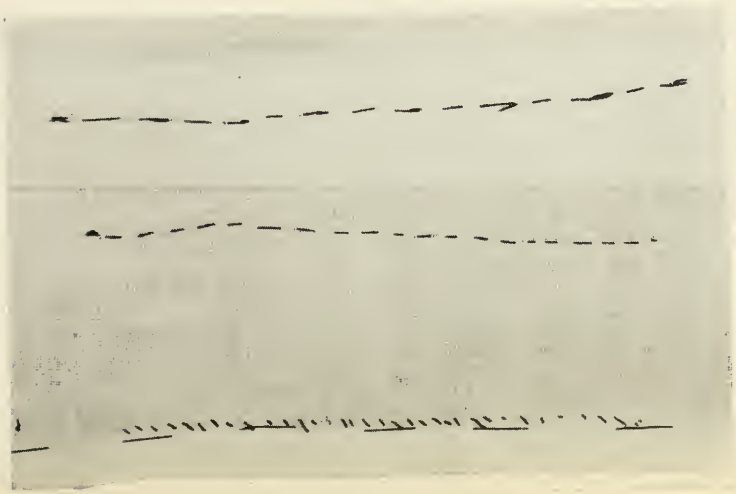


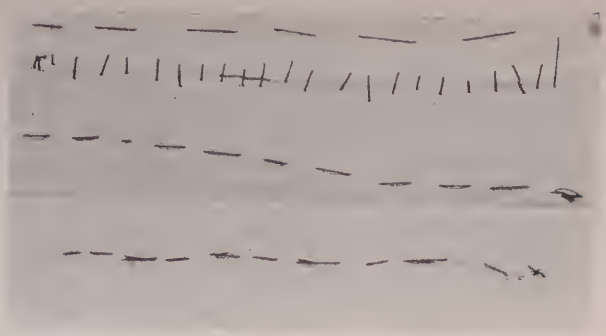
4.3

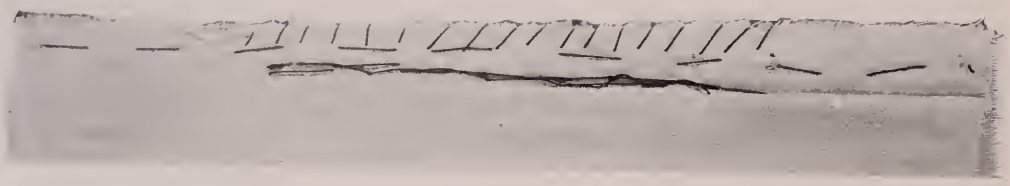


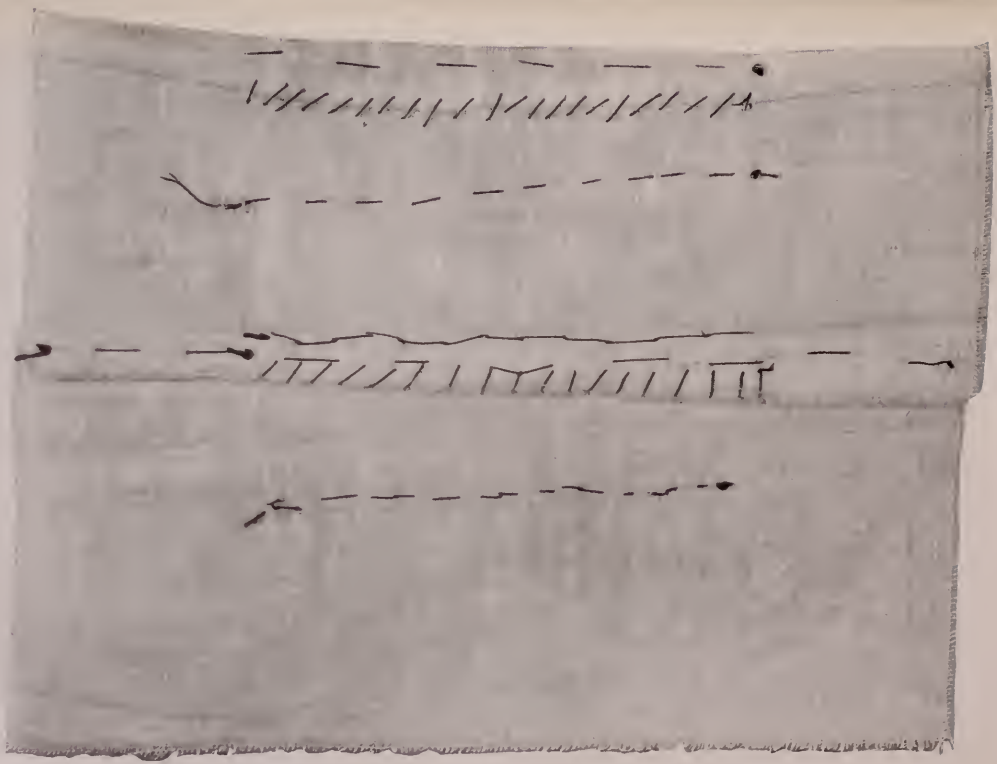
6.1



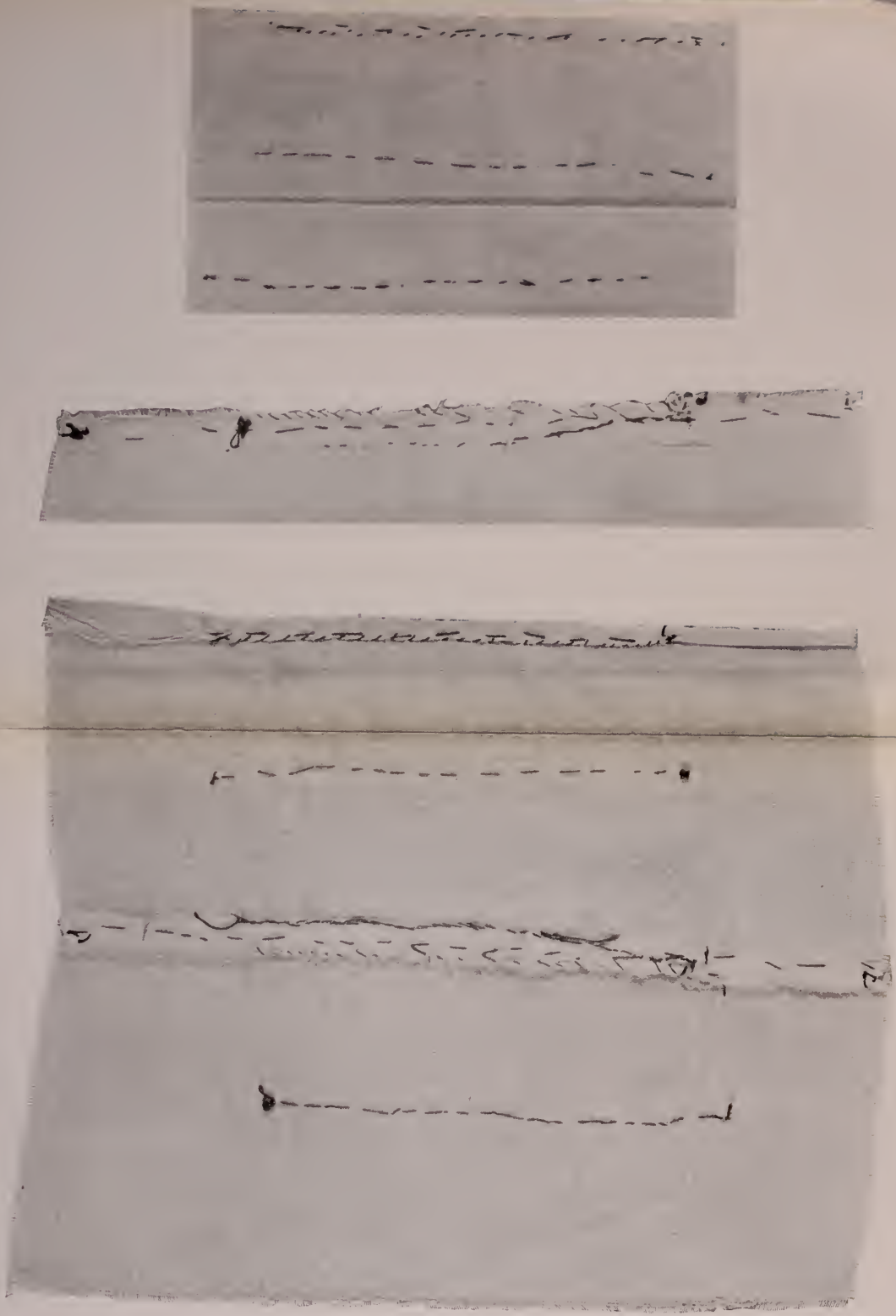




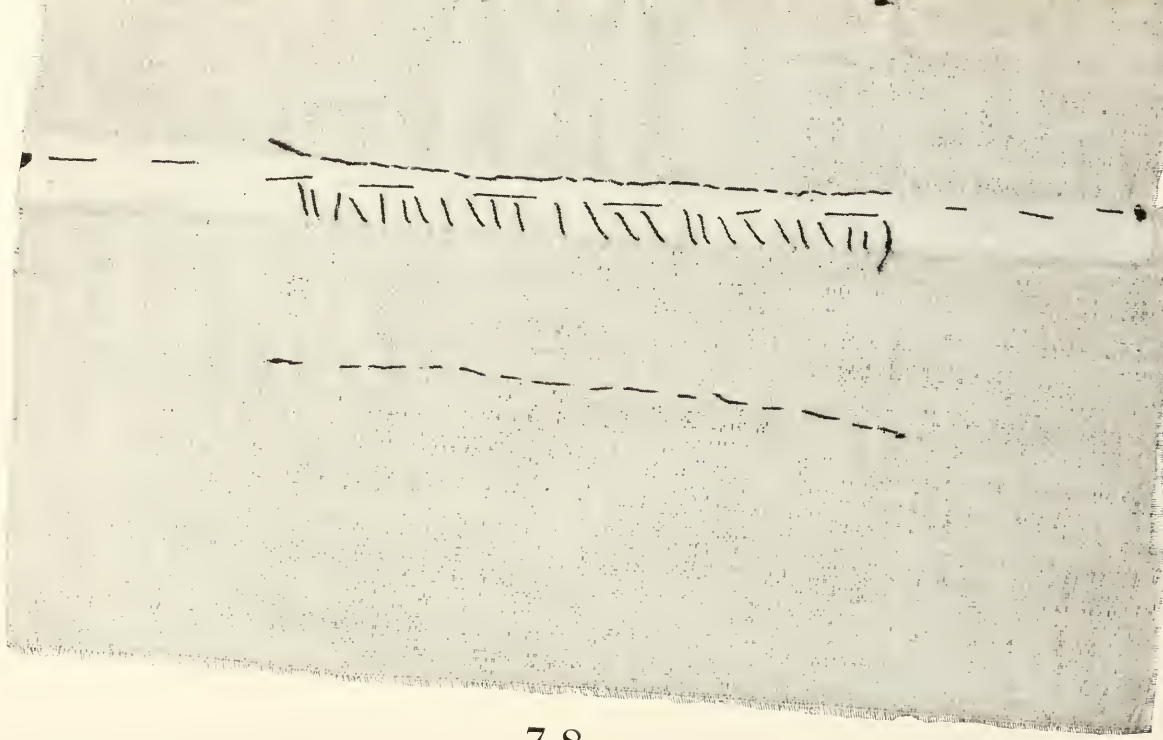




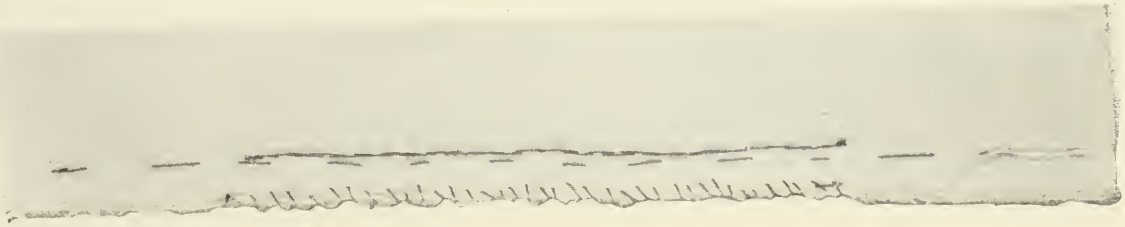
5.1

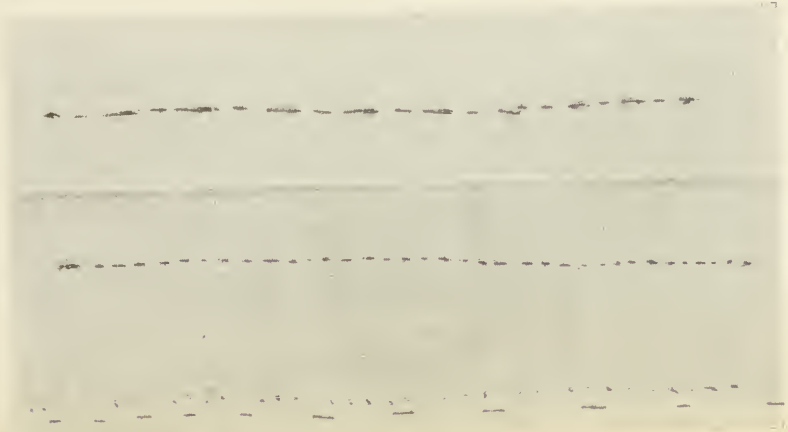


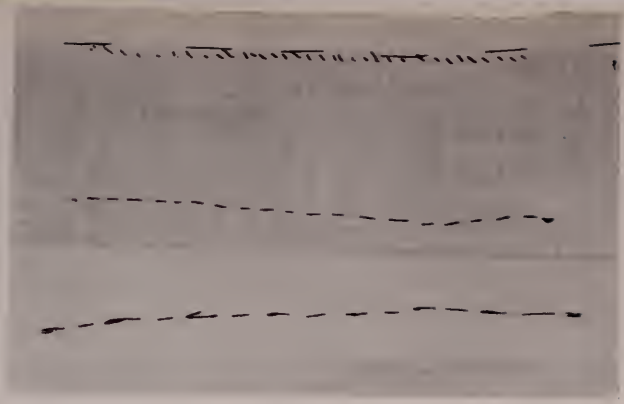
6.1

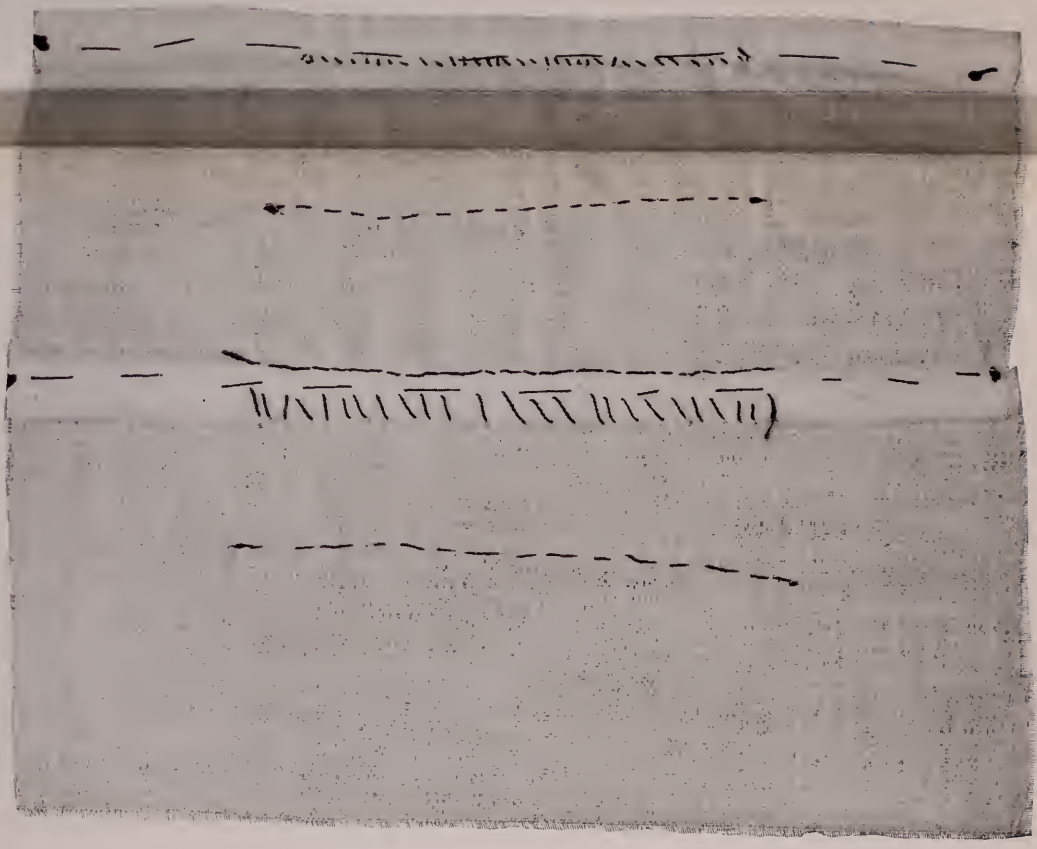


7.8

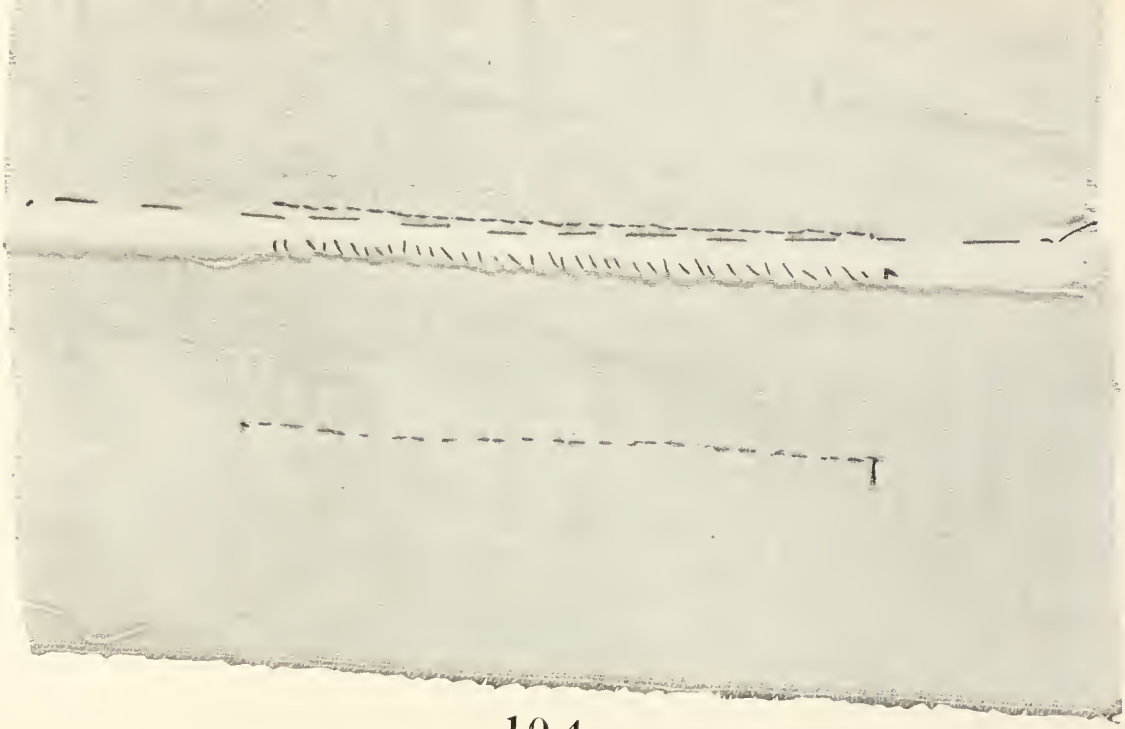




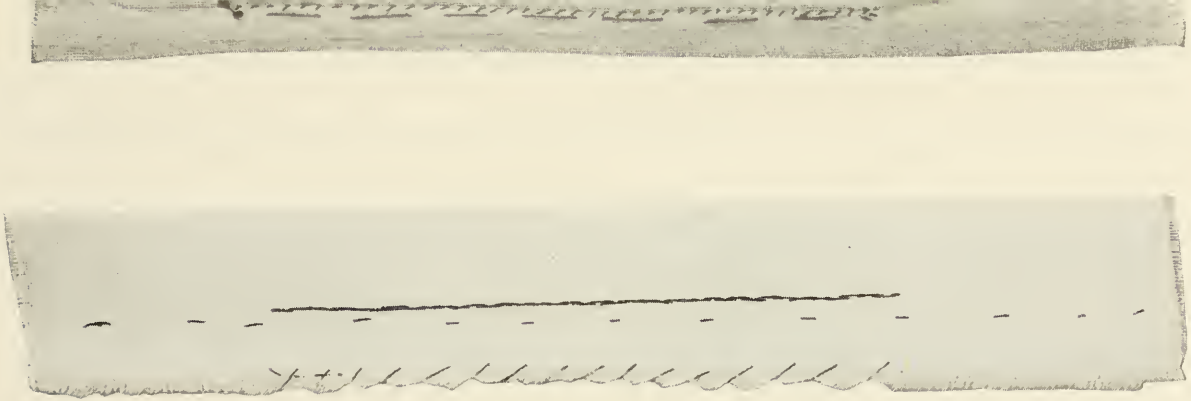


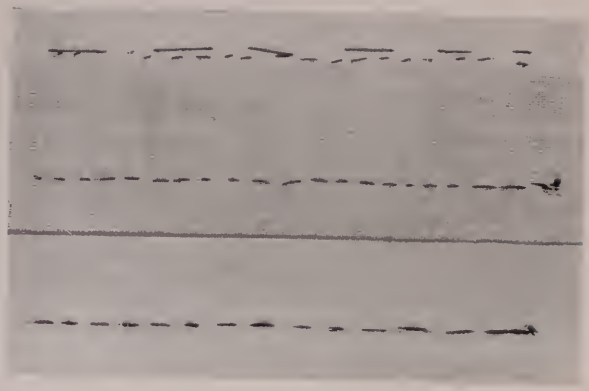


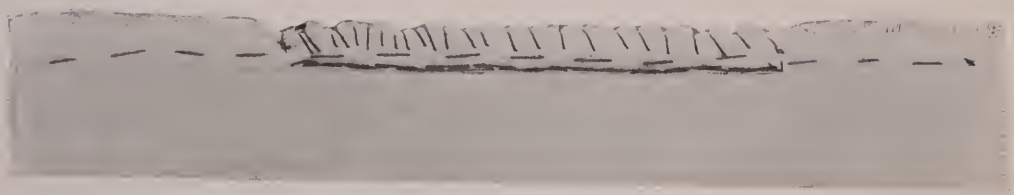
7.8

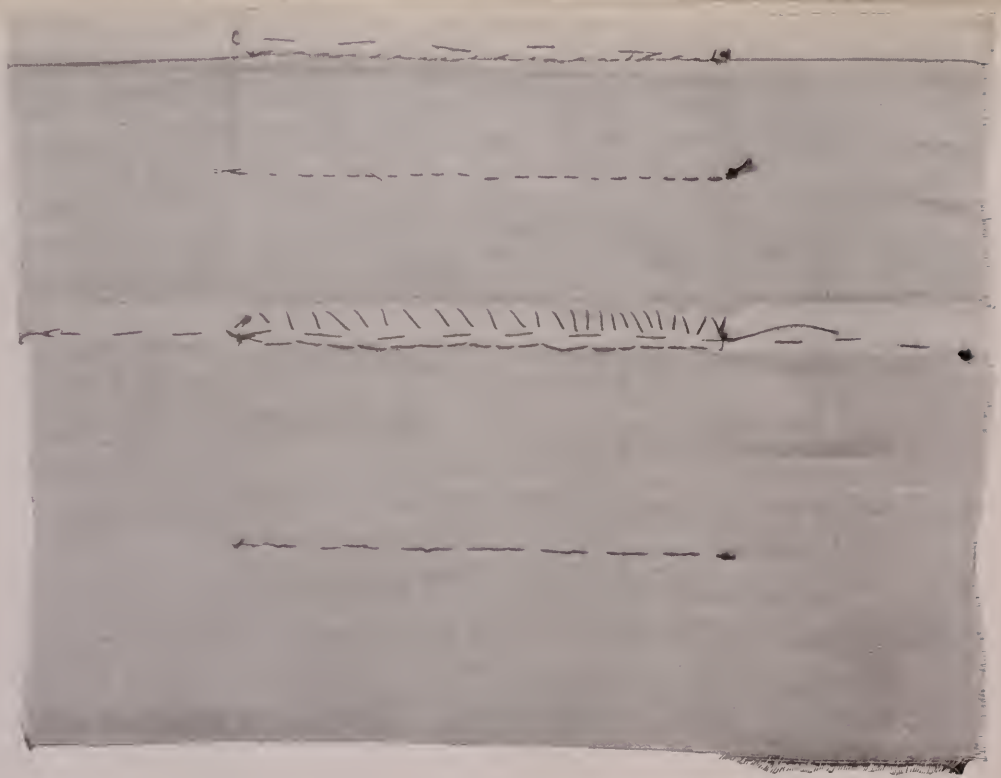


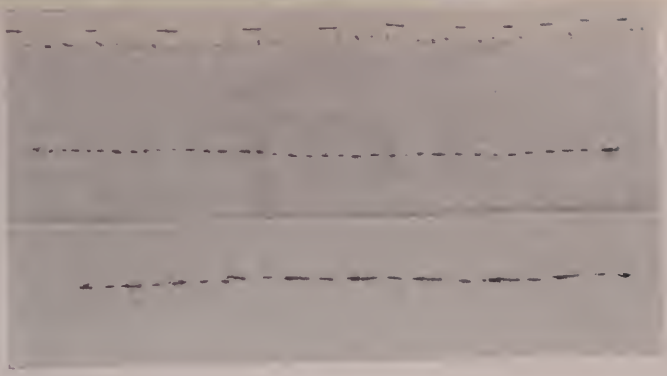
10.4

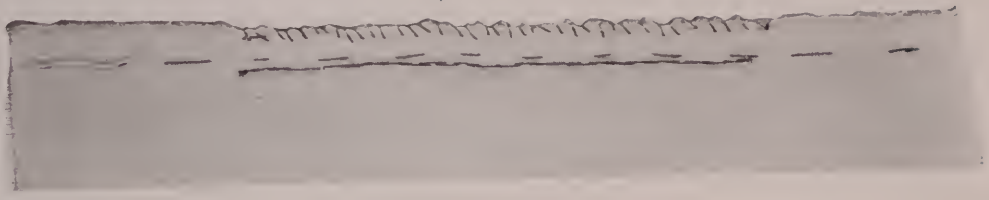


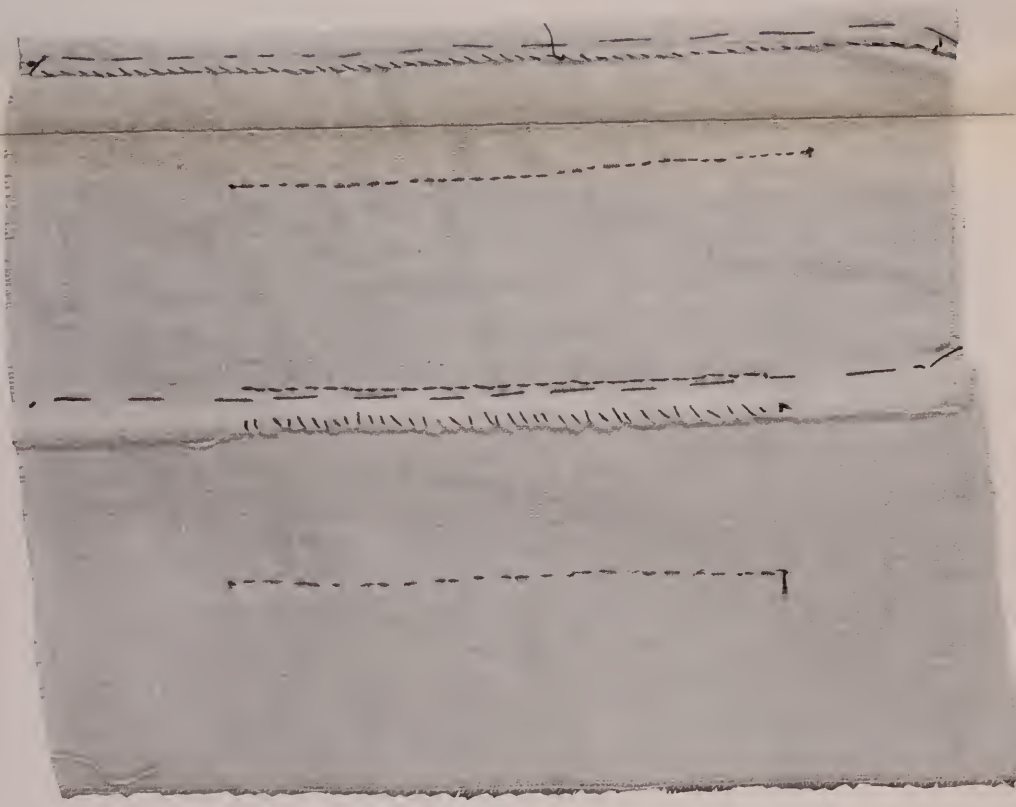




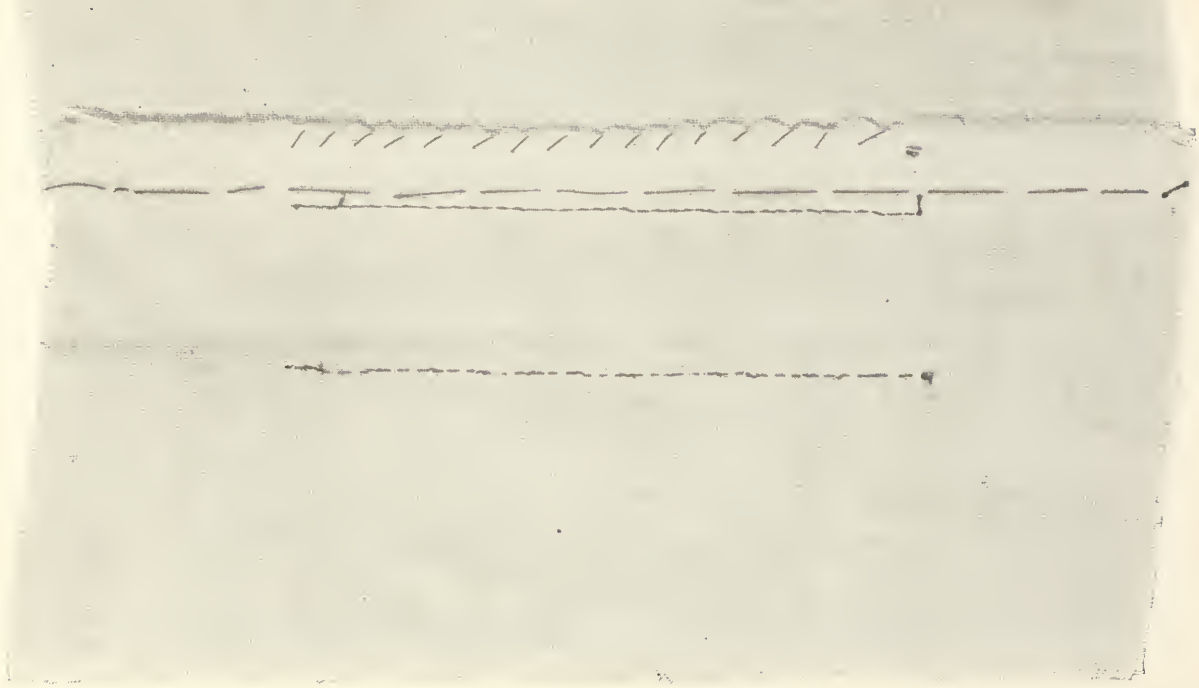




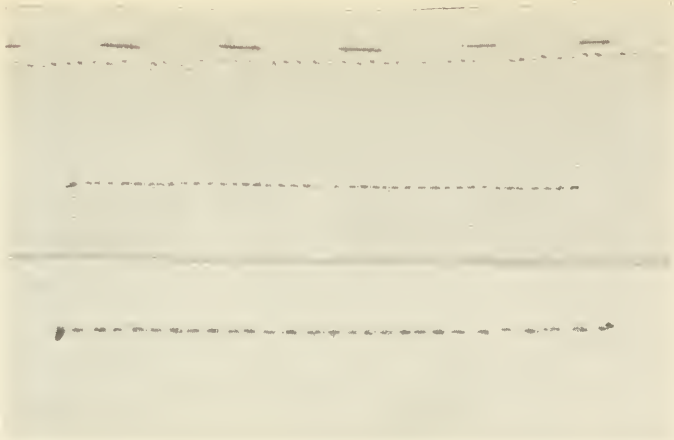


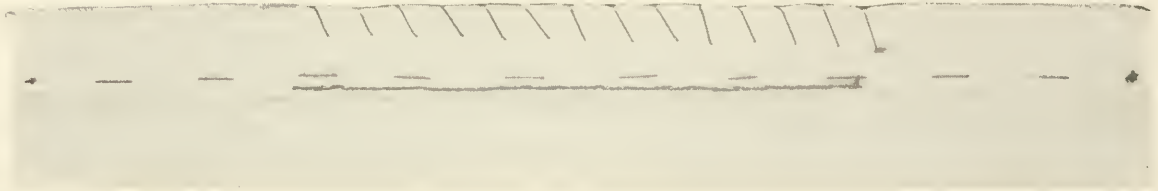


10.4

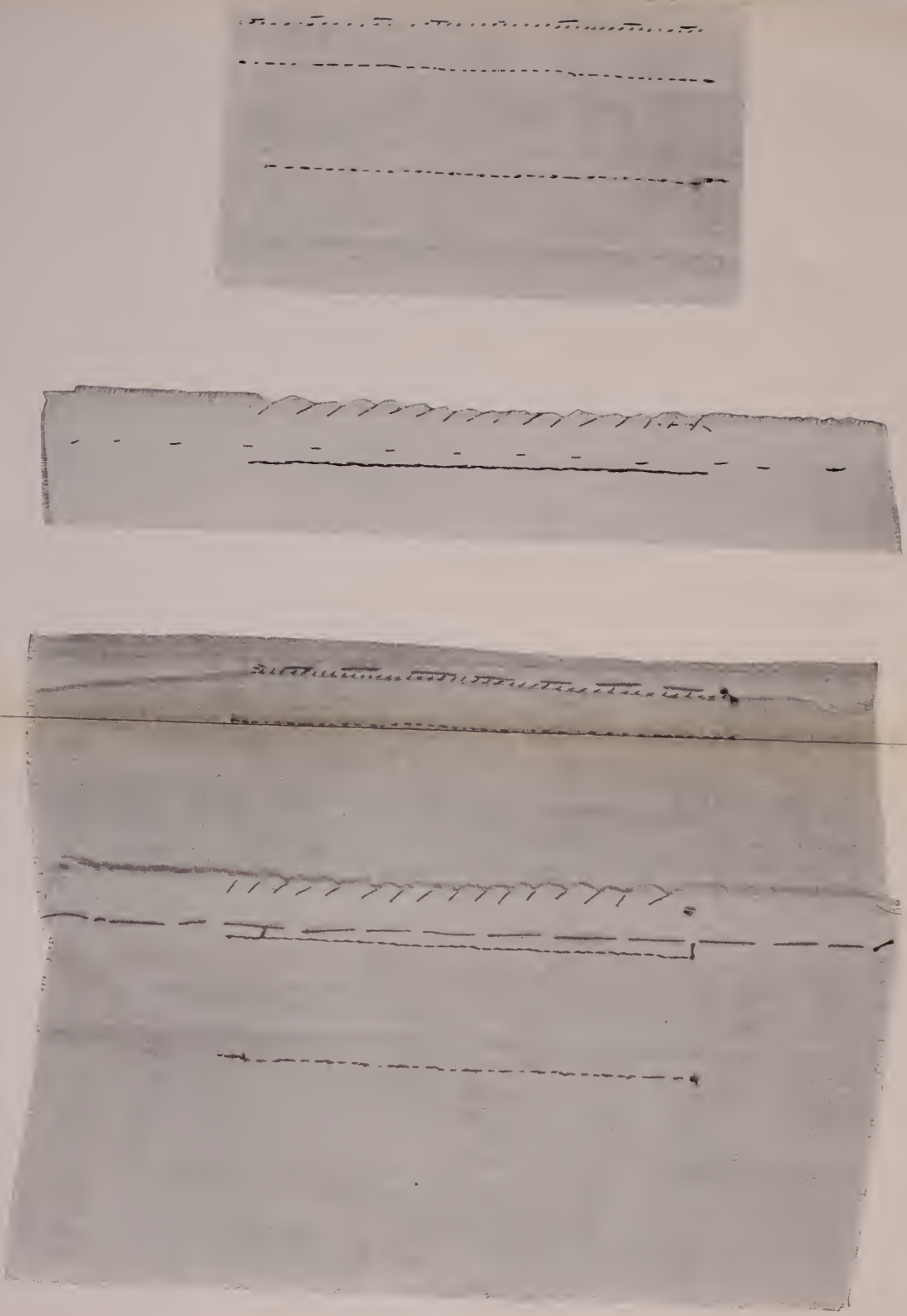


13.1



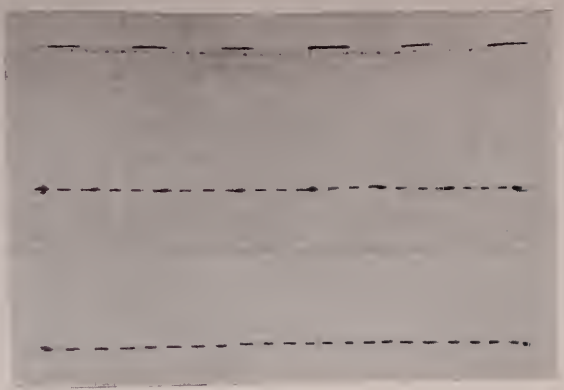


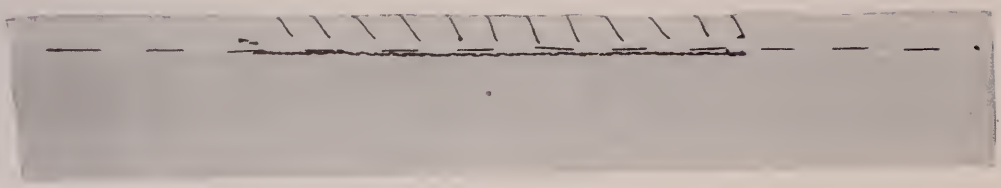
11.5

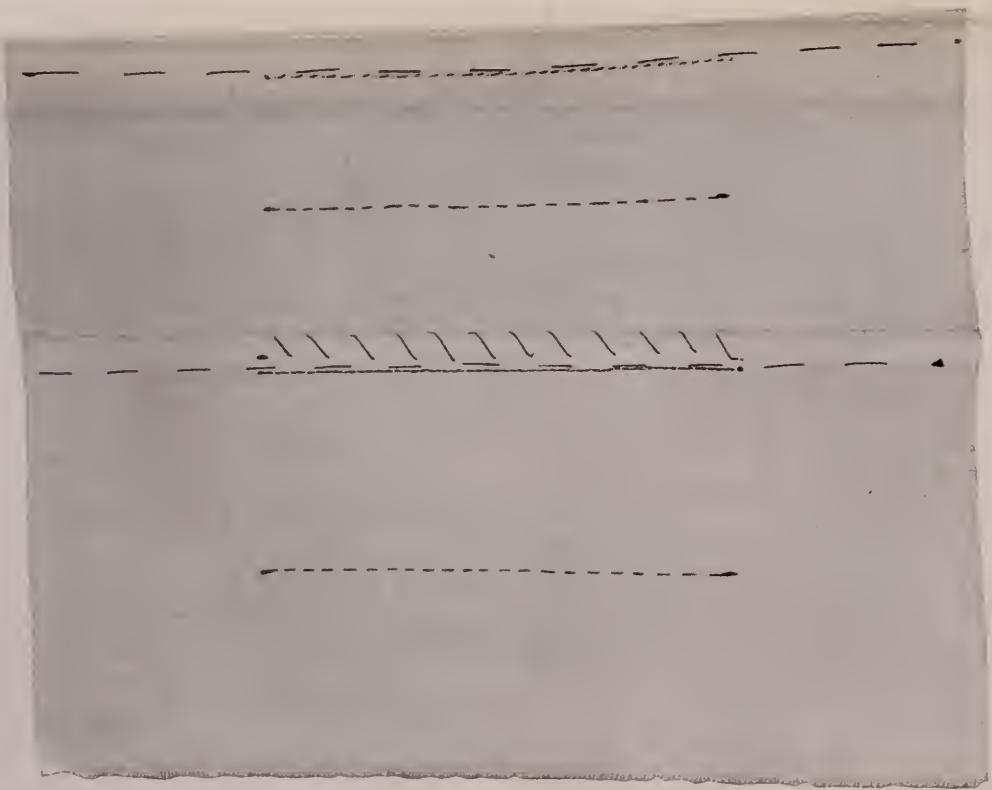


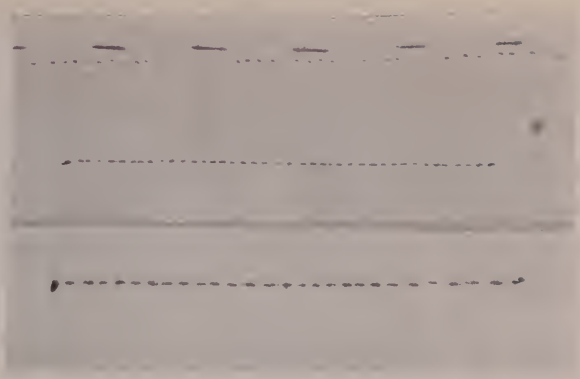
13.1

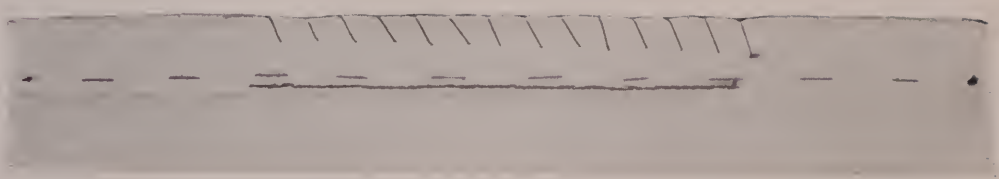
15.1

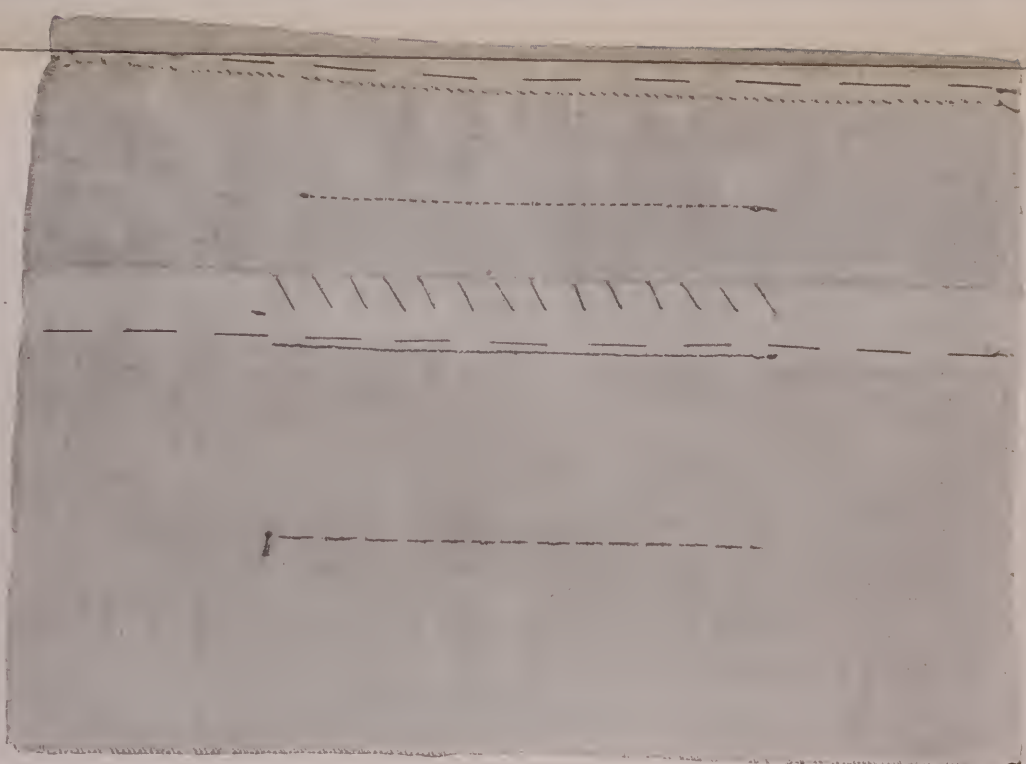




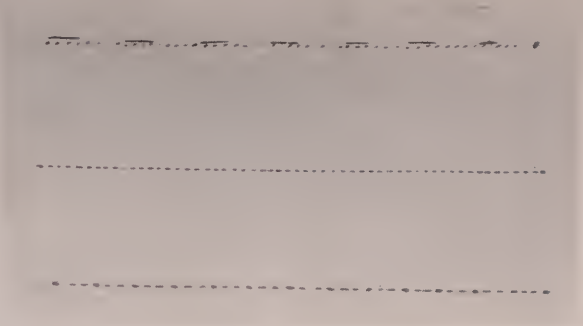


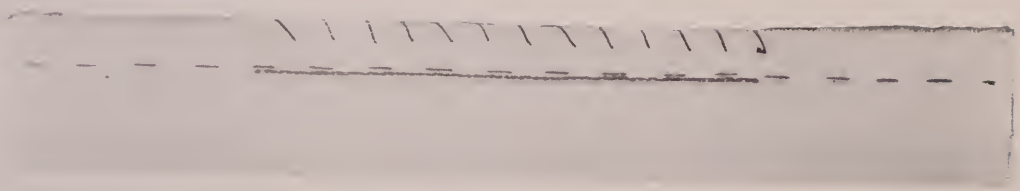


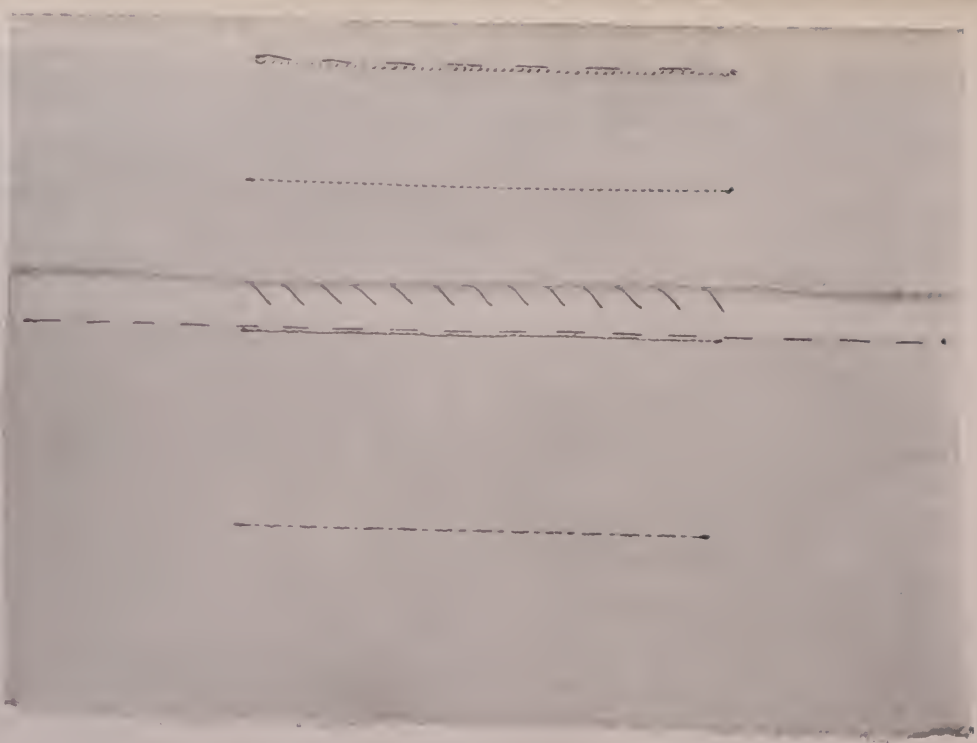




15.1







LIBRARY OF CONGRESS
0 014 082 508 6